AN OBJECTIVE CLASSIFICATION OF VOICE DISORDERS USING ARTIFICIAL NEURAL NETWORKS.

Binu (R)

Reg. No. M9601

A DISSERTATION SUBMITTED AS PART FULFILMENT OF FINAL YEAR M.Sc. (Speech and Hearing)

All India Institute of Speech & Hearing
Mysore.

May 1998

DEDICATED TO

"HARI"

My life, my energy, my strength, my courage

Hold me, support me and be with me showing me the right way as you have always done.

CERTIFICATE

This is to Certify that the dissertation entitled "AN OBJECTIVE CLASSIFICATION OF VOICE DISORDERS USING ARTIFICIAL NEURAL NETWORKS'" is the bonafide work in part fulfilment for the degree of Master of Science (Speech and Hearing) of the student with Register No. M9601.

Mysore
May, 1998

Dr. (Miss) S. NIKAM
Director
All India Institute of Speech & Hearing
MYSORE - 570 006

CERTIFICATE

This is to Certify that the dissertation entitled "AN OBJECTIVE CLASSIFICATION OF VOICE DISORDERS USING ARTIFICIAL NEURAL NETWORKS" has been prepared under my supervision and guidance.

Mysore

May, 1998

Dr. N.P. NATARAJA
Professor and H.O.D
All India Institute of Speech & Hearing
MYSORE - 570 006

DECLARATION

This dissertation entitled "AN OBJECTIVE CLASSIFICATION OF VOICE DISORDERS USING ARTIFICIAL NEURAL NETWORKS" is the result of my own study under the guidance of Dr. N. P. NATARAJA, Professor and H.O.D. Department of Speech Sciences, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier at any University for any other diploma or degree.

Mysore　　　　Reg. No.M9601

May, 1998

ACKNOWLEDGEMENTS

I am extremely grateful to my teacher and guide, **Dr. N.P. NATARAJA,** Professor and H.O.D. Department of Speech Sciences, All India Institute of Speech and Hearing, Mysore, for his constant support, encouragement and guidance for completing the dissertation.

I would like to thank **Dr. (Miss) S.NIKAM,** Director, All India Institute of Speech and Hearing, Mysore, for permitting me to take up this topic for my dissertation.

My grateful thanks to **Mrs. SREEDEVI and Miss LALITHA** for helping me to sort out my "Computer problems".

Dad, Mom and **Bro,** you mean a lot to me

Deep, Sara, Manoj, Sajith, JP, Veera, Ratna, Prashanth and **Riza,** thanks for being such good friends and making my life exiting in AIISH.

Duck, my **cute little** sister, thank you for your advices encouragement and support.

Turtle, thank you for being such a lovely friend you have made me realise what a 'friend' means. Here is wishing you all the best in all your future endeavours.

Dear **Thambi's - Muthu, Brito, Abbas, Chaiidau, Milind, Kathir** and **Gokul,** an informal thanks to you all in giving me a helping hand whenever needed, entertaining me and making my stay at AIISH memorable.

Hats off to you junior guys and girls **- Srivatsan, Siddharth, Amith, Mukundan**, **Sharad** and **Kavitha.** Thank you for having helped me in my work.

Last but not the least **'Neural connection'** - thank you for taking me into the new horizons of advanced computer technology and teaching me the basics of my study.

CONTENTS

PAGE NO.

LIST OF TABLES

LIST OF GRAPHS

LIST OF FIGURES

INTRODUCTION

Voice has been defined as "the laryngeal modulation of the pulmonary air stream, which is further modified by the configuration of the vocal tract" (Micheal and Wendahl, 1971). The production of voice depends upon the synchrony or coordination between respiratory, phonatory and resonatory systems. Deviation in any of these systems may leads to voice problems.

Fant (1960) defines voice by using the formula P = ST, where "P" the speech found is the product of source 'S' and the transfer function of the vocal tract 'T'. An attempt has been made by Nataraja and Jayaram (1975) to review the definitions of the normal voice critically. They conclude that each of the available definitions of voice have used subjective terms, which are neither defined or measurable. They have suggested a possibility of defining good voice operationally by stating that "the good voice is one which has optimum frequency as its fundamental (habitual) frequency".

The production of voice is a complex process. It depends on the synchrony between respiratory, phonatory and resonatory systems which in turn requires precise control by the central nervous system.

Hirano (1981) states that, "during speech and singing the higher order centres in the cerebral cortex control voice production and all activities of the central nervous system is finally reflected in the muscular activity of the voice organs". Because of the interdependence of the respiratory, phonatory and resonatory systems during voice production disturbances in any one of the systems may lead to deviant or abnormal voice quality. Voice plays a major role in speech and hence communication.

The ultimate aim of the studies on normality and abnormality of voice and assessment and diagnosis of voice disorders is to enforce a procedure which will eventually bring back the voice of an individual to normal or optimal level.

Many have suggested various means of analyzing voice to note the factors that are responsible for creating an impression of a particular voice and to determine the underlying mechanism (Michael and Wendahl, 1971; Jayaram, 1975; Fritzell and Hammerberg, 1977; Hirano, 1981; Nataraja, 1972; Rashmi, 1985).

There are objective methods like Electro-myogaraphy (EMG), stroboscopy, ultrasound, glottography, ultra high speed, photography, photoelectric glottography, electroglottography, aerodynamic measurements, acoustic analysis, etc. which measure varous aspects of voice Presently acoustic and aerodynamic analysis of voice is gaining more importance. Studies have considered in past the effectiveness of various acoustic and aerodynamic parameters of voice in differentiating normals from dysphonics (Jayaram, 1975; Nataraja, 1986; Hirano, 1981) and also monitoring pre- and post-treatment changes in voice (Cooper, 1974; Vanderburg and Hocksema, 1980; Wedin and Orgen, 1982; Trullinzer and Emanual, 1988; Hufnagle and Hufnagle, 1989; Susheela, 1989; Schulte, Kitzury and Akerfund, 1993; Menon, 1996). The parameters studied have varied frmo investigator to investor.

Michael and Wendahl (1971) consider voice as a series of measurable events and suggest twelve parameters for assessing the voice and voice disorders. Nataraja (1986) studied 22 acoustic and aerodynamic parameters in normals and dysphonics. He concluded that only six parameters were sufficient to differentially diagnose dysphonics from normals.

After the measurement of parameters, the clinician has to make a judgement regarding the diagnosis based on the values of parametrs. Since the computers have been put into use in almost all spheres of life, attempts have been made to use them for classifying the disorders, based on the symptoms and values of parameters measured. An Artificial Neural Network is a computer branch which has been widely applied for such activities.

Artificial Neural Networks (ANNs) are 'biologically' inspired networks. They have ability to learn from empirical data or information (Raol and Mankame, 1996).

ANNs have apparent ability to imitate the brain's activity to make decisions and draw conclusions when presented with complex and noisy informations (Raol and Mankane, 1996).

Neural networks are simply a new way of analyzing data. The revolutionary aspect of neural networks which makes them so unique is their ability to learn complex patterns and trends in data. Neural networks acquire knowledge by training a set of data. After the network has been trained and validated, the model may be applied to data it has not seen previously for prediction classification, time series analysis or data segmentation.

Studies using neural networks in the field of speech and hearing are scanty and ANNs applied to the field of voice and voice disorders are still less.

Few studies have used self organised Map (SOM) which is an algorithm of Kohenen network to classify normal and dysphonic voices to find objective indices for dysphonic voice (Leinonen, Kangas, Torkkola and Juvas, 1992), to find acoustic pattern recognition of fricative-vowel coarticulation by self organizing map (Leinonen, Mujunen, Kangas, Torkkola and Juvas, 1993), acoustic pattern recognition of /s/ misarticulation by SOM (Mujunen, Leinonen, Kangas and Torkkola, 1993), model for control of voice fundamental frequency by a neural network (Farley, 1994) and cotegorization of voice disorders on perceptual dimensions (Leinonen, Kangas, Torkkola and Mujunen, 1997).

Hence the present study attempts to classify voice disorders in terms of its acoustic and aerodynamic parameters by using a neural network.

The parameters used for the study to assess voice have been suggested by Nataraja (1986). They are

Aerodynamic parameters

1. Mean air flow rate
2. Maximum duration of phonation

Acoustic parameters

1. Fundamental frequency in phonation
2. Extent of fluctuations in fundamental frequency
3. Speed of fluctuations in fundamental frequency

4. Extent of fluctuations in intensity
5. Speed of fluctuations in intensity
6. Frequency range in phonation
7. Intensity range in phonation
8. Fundamental frequency range in speech.

These aerodynamic and acoustic parameters were fed into a neural network. The neural network used for the present study was Multi Layer Perception (MLP).

AIM OF THE STUDY

1. Hypothesis

a. Main hypothesis: There is no significant difference between normals and dysphonics both males and females for the aerodynamic and acoustic parameters studied.

b. Auxiliary hypothesis:

i. There is no significant difference between normal males and normal females for the aerodynamic parameters studied.

ii. There is no significant difference between dysphonic males and dysphonic females for the aerodynamic and acoustic parameters studied.

The aerodynamic and acoustic parameters studied are:

1. Aerodynamic parameters
    - i. Mean air flow rate (MAFR)
    - ii. Maximum phonation duration (MPD)

2. Acoustic parameters
    i. Fundamental frequency in phonation
    ii. Extent of fluctuations in frequency
    iii. Speed of fluctuations in frequency
    iv. Extent of fluctuations in intensity
    v. Speed of fluctuations in intensity
    vi. Frequency range in phonation
    vii. Intensity range in phonation
    viii. Fundamental frequency range in speech.

2. To obtain a model for objective classification of different voice disorders based on aerodynamic and acoustic parameters using a neural network.

3. To use this model created by the neural network for further classifying the voice disorders using the acoustic and aerodynamic parameters.

IMPLICATIONS OF THE STUDY

1. The study used a neural network which has a capability to learn and understand complex data and hence creates an objective classification model for voice disorders.

2. Such applications can be attempted into other speech and hearing disorders. This can be used for a regular clinical activity.

LIMITATION

1. The number of subjects in the study were limited.

2. The number of subjects were not uniformly distributed in terms of age and sex.

REVIEW OF LITERATURE

"The act of speaking is a specialised way of using the vocal mechanism. The act of singing is even more so. Speaking and singing demand a combination or interaction of the mechanisms of respiration, phonation, resonation and articulation" (Boone, 1983). The underlying basis of speech is voice. The importance of voice in speech is very well depicted when once considers the cases of laryngectomy or even voice disorders.

"Voice plays the musical accompaniment to speech, rendering it tuneful, pleasing, audible and coherent, and is an essential feature of efficient communication by the spoken work" (Green, 1964).

It has been well established that voice has both linguistic and non-linguistic functions in any language.

Voice is the carrier of speech, variations in voice in terms of pitch and loudness, provide rhythm and also break the monotony. This function of voice draw attention when there is a disorder of voice.

"Voicing (presence of voice) has been found to be a major distinctive feature in almost all languages. 'Voicing' provides more phonemes and makes the language broader. When this function is 'absent' or used 'abnormally' it would lead to a 'speech disorder'.

At the semantic level also voice plays an important role. The use of different pitches, high and low, with the same string of phonemes would mean different things. Speech prosody - the tone, the intonation and the stress or the rhythm of language is a function of vocal pitch and loudness as well as of phonetic duration.

Perkins (1971) has identified atleast five non-linguistic functions of voice. Voice can reveal speaker identity, i.e. voice can give information regarding sex, age, height and weight of the speaker. Lass, Brong, Ciccolella, Walters and Maxwell (1980) report several studies which have shown that it was possible to identify the speaker's age, sex, race, socio-economic status, racial features, height and weight based on voice.

It is a prevailing notion that there is a relationship between voice and personality, i.e. voice reflects the personality of the individual (Stark Weather, 1961 and Markel, Meisels and Hauck, 1964).

Voice has also been considered reflecting the physiological state of an individual.

The basic parameters of phonation are

1. Parameters which regulate the vibratory pattern of the vocal folds.

2. The parameters which specify the vibratory pattern of vocal folds.

3. The parameters which specify the nature of sound generated (Cotz, 1961).

Hirano (1981) has further elaborated on this, by stating that "the parameters which regulate the vibratory pattern of vocal folds can be divided into two groups - physiological and physical. The physiological factors are those related to the activity of the respiratory, phonatory and articulatory muscles. The physical factors include the expiratory force, the conditions of the vocal folds and the state of the vocal tract.

The vibratory pattern of the vocal folds can be described with respect to various parameters including the fundamental frequency, regularity or periodicity in successive vibrations, symmetry between the two vocal folds, uniformity in the movement at different points within each vocal fold, glottal closure during vibration, contact between the two vocal folds and so on.

The nature of sound generated is chiefly determined by the vibratory pattern of the vocal folds. It can be specified both in acoustic terms and in psychoacoustic terms. The psycho-acoustic parameters are fundamental frequency, intensity, acoustic spectrum, and their time-related variations. The psycho-acoustic parameters are pitch, loudness and quality of voice and their time related changes.

Acoustic analysis has been considered as the basic tool in the investigation of voice disorders. It has been considered vital in the diagnosis and the management of patients with voice disorders. Hirano (1981) has pointed out that the acoustic analysis of voice signal may be one of the most attractive method for assessing phonatory functions or laryngeal pathology because it is non-invasive and provides objective and qualitative data.

Further, a clinician will not really know what to expect with a medical diagnosis having a description of the larynx together with some adjectives like 'hoarse or rough' until he actually sees the case (Michael and Wendahl, 1971). On the other hand if the clinician gives a report which includes measures of frequency ranges, respiratory function, jitter, shimmer their related variations, noise and harmonic components, etc. in the form of a voice profile, the clinician can then compare these values to the norms for each one of the parameter and thus have a relatively good idea as to how to proceed with therapy even before seeing the patient. Moreover, periodic measurement of these parameters during the course of therapy may well provide an useful index so as to the success of the treatment (Michael and Windahl, 1971).

Deliyski (1990) presented an acoustic model of pathological voice production which described the non-linear effects occurring in the acoustic wave form of disordered voices. The voice components such as fundamental frequencies and amplitude irregularities and variations. Sub-harmonic components, turbulent noise and voice breaks are formally expressed as a result of random time function influence on the excitation function and the glottal jitter. Quantitative evaluation of these random functions is done by computation of their statistical characteristics which can be used in assessing voice in clinical practice. This set of parameters, which corresponds to the model, allows a multi-dimensional voice quality assessment. Since any single acoustic parameter is not sufficient to demonstrate the entire spectrum of vocal function or of laryngeal pathology, multidimensional analysis using multiple acoustic parameter has been attempted by some investigators. Davis (1976) used parameters such as pitch perturbation quotient, amplitude perturbation quotient, pitch amplitude, coefficient of excess, spectral flatness of inverse filter spectrum and spectral flatness of the residue signal spectrum and performed multidimensional analysis at differentiation of pathological voices. The detection was 95.2% in a closed test and 67.4% in an open test.

Hirano (1989) who carried out an international survey has found the following measures as being used for clinical voice evaluation:

1. Airflow
   Phonation Quotient (PQ)
   Vocal Velocity Index (VVI)
   Maximum Phonation Time (MPT)
2. F_0 range
   SPL range
   Habitual F_0
   Habitual SPL
3. Electroglottography
4. Tape recording
   Pitch perturbation
   Amplitude perturbation
   S/N ratio
   LTAS
   Inverse filter acoustic
   VOT
   Perceptual evaluation

5. Laryngeal mirror
   Fibroscopy of larynx
   Microscopy of larynx
6. X-ray laryngography
7. Vital capacity
   Ribcage movements
8. Audiometry

There are various objective measures to evaluate these parameters. Stroboscopic procedure, pitchmeter, high speech cinematography, electroglottography, digi pitch, pitch computer, ultrasonic recordings and the high resolution signal analyzer. At present various computer based methods are being evolved which are fast in terms of analysing the voice samples and giving the values of the parameters as such. Recently these methods are being used mostly in clinical and research work because they are time saving and they don't need interpretation on the part of the experimenter since the parameters are automatically analyzed and the values are given.

Micheal and Wendahl (1971) consider voice as a multi-dimensional series of measurable events, implying that a single phonation can be assessed in different ways. They present a tentative list of twelve parameters of voice, "most of which can be measured and correlated with specific perceptions, while others are more elusive and difficult to talk about in more than ordinal terms. The twelve parameters listed by them are:

1. Vital capacity
2. Maximum duration of controlled sustained phonation
3. Modal frequency range
4. Maximum frequency range
5. Maximum duration of sustained parameters
6. Volume/velocity of airflow during phonation
7. Glottal wave form
8. Sound pressure level
9. Jitter of the vocal signal
10. Shimmer of the vocal signal
11. Effort level (vocal)
12. Transfer function of the vocal tract

Jayaram (1975) made an attempt to develop a method of differential diagnosis of dysphonia based on the measurement of the following parameters in normals and dysphonics.

a. Optimum frequency
b. Habitual frequency
c. Frequency range
d. Maximum phonation duration
e. Vital capacity

f. Mean airflow rate
g. Vocal velocity index

From this study he concluded that these parameters were useful in differentiating dysphonics from normals and further in differentiating different types of dysphonics. Several other studies also related different parameters to different dysphonics.

Nataraja (1986) used 23 acoustic aerodynamic and spectral parameters in an attempt to differentially diagnose different dysphonics. The parameters are enlisted below:

1. Vital capacity
2. Mean air flow rate
3. Phonation quotient
4. Vocal velocity index
5. Maximum phonation duration
6. S/2 ratio
7. Fundamental frequency in phonation
8. Fundamental frequency in speech
9. Optimum frequency
10. Extent of fluctuation in fundamental frequency
11. Extent of fluctuation in intensity
12. Speed of fluctuation in fundamental frequency
13. Speech of fluctuation in intensity
14. Frequency range in phonation

15. Frequency range in speech
16. Intensity range in phonation
17. Intensity range in speech
18. Rising time in phonation
19. Falling time in phonation
20. Fundamental frequency
21. Ratio of intensities about 1 KHz and below 2 KHz.
22. Ratio of intensities of harmonics and noise in 2-3 KHz.
23. Frequency of first formant.

He concluded that six parameters were sufficient to differentiate the dysphonics from normals. They were

1. Speed of fluctuations in fundamental frequency
2. Maximum phonation duration
3. Frequency range in phonation
4. Falling time in phonation
5. Extent of fluctuations in frequency
6. Ratio of intensities above and below 1 KHz (alpha-ratio)

And eight parameters were sufficient to differentially diagnose voice disorder. They are

1. Mean air flow rate
2. Phonation quotient
3. Vocal velocity index
4. Maximum phonation duration

5. Fundamental frequency in phonation
6. Speed of fluctuations in fundamental frequency
7. Extent of fluctuations in intensity
8. Speed of fluctuations in intensity
9. Frequency range in phonation
10. Intensity range in phonation
11. Rising time in phonation
12. Falling time in phonation
13. Fundamental frequency in speech
14. AA ratio

Traditionally, air flow measures, subglottal pressure maximum phonation, mean airflow rate, glottal resistance, vocal efficiency had been measured (Isshiki and Von Leden, 1964; Yanajihara and Von Leden, 1967; Hirano, Koike and Von Leden, 1968; Devata, Von Leden and Williams, 1972; Smith et al., 1992; Holmbey et al., 1994). These parameters help in delineating the interaction of the respiratory system and the laryngeal system to produce perfect modulation of the air stream at the glottis. The smooth flow of air is later modulated by the upper airway dynamics. There have been studies regarding the subglottal pressure, airflow rate, etc. in normals and dysphonics in various pathologies like tumour, paralysis, contact ulcer, edema, etc. using various instruments like spirometer and pneumatachograph, etc. (Isshiki and Von Leden, 1964; Isshiki et al.f 1967; Hirano et al., 1968; Yoshioka et al., 1977; Shigemori, 1977).

The presence of a laryngeal disorder produces an airflow waveform that shows an increase in turbulence. This variation in airflow waveform is attributed to the loss of the vocal folds ability to sustain periodic vibration. Thus the aerodynamic measures have been found to be a clinically useful measures for analysing vocal dysfunction and have led to the better understanding of laryngeal disorders.

Vital capacity

The maximum volume of air that can be expired after a deep inhalation is termed as vital capacity, which is measured in terms of cc.

The amount of air available for an individual for an individual for the purpose of voice production depends on the vital capacity of an individual. High lung volume helps in sustaining the voice/speech for a longer duration (Bouheys, Proctor and Mead, 1966). The measurement of vital capacity is important as it provides an estimate of the amount of air potentially available for the production of voice. The mechanical function of lungs as an air (power) supply source has been tested through the measurement of vital capacity.

Various factors have been reported to be affecting the vital capacity. Vareed (1981) has reported that the vital capacity varies with age, sex, height, weight, body surface area, body build, the amount of exercise and other factors. Hutchinson (1948) has demonstrated that the vital capacity and body size are correlated with arithmetical progression and that the age and the weight seem to be significant only in extreme cases of variation, the circumference of the chest having no immediate influence on the vital capacity.

Krishna and Vareed (1932) studied 103 males, age ranging from 18-29 years, from South India to obtain vital capacity, weight and height, body surface area and chest circumference. They reported that the average vital capacity to be low (2.93 litres) compared to North Indians. They attribute this low vital capacity not to race but to the warm climate, less tendency for excessive low metabolism and poor chest expansion. Nag, Chatterjee and Dey (1982) have reported that the lung function consistantly declined with age. Mandi, Lengyal and Csukas (1976) established that the decrease in vital capacity was more definite in old age. Similar reports have been made by Meenakshi (1984) from a study of 60-80 year old normal subjects.

Males have shown higher values of vital capacity than females (Jain and Ramaiah, 1969; Jayaram, 1975;

Nataraja and Rashmi, 1984 and Krishnamurthy, 1986). Verma et al. (1983) report that mean vital capacity values in Indians were significantly lower than in the Western subjects.

Nadoleczny and Luchsinger (1934) noticed significantly higher vital capacity values in well trained athletes and professional singers. Sheela (1974) reported that there was no significant difference in vital capacity between the trained and untrained singers (both in males and females). The mean vital capacity for males and for females in the age range of 19-54 years were 2685 cc and 1574 cc, respectievly.

Jayaram (1975) reported a mean vital capacity value of 3180 cc and 2210 cc in normal adult males and females, respectively. Nataraja and Rashmi (1985) have reported the mean vital capacity for adult males and females as 2950 cc and 1750 cc, respectively.

Yanagihara and Koike (1967) have related vital capacity to phonation volume; while Hirano, Koike and Von Leden (1968) found a relationship between vital capacity and maximum phonation duration. Yonagihara and Koike (1960) have reported that both the phonation volume and the ratio of phonation volume to vital capacity, decrease as the pitch level decreases. A correlation between vital capacity and phonation volume was reported with coefficients ranging from

0.59 to 0.90. Hirano et al. (1968) have determined the ratio of vital capacity and maximum phonation duration and have termed it 'phonation quotient'. Higher flow rates in normal subjects were generally associated with shorter phonation durations or lower vital capacities. Bouhuys et al. (1966) have shown that the singers designated as having 'poor quality' had smaller vital capacities than singers categorized as having 'good' or 'average' quality.

Darby (1981) while discussing the interaction between speech and disease, states that "many respiratory disorders of mild severity and slow onset may not interfere substantially with speech production because they can be partially compensated for by gradual adjustments in respiratory behaviour and modifications in speaking habits. A small reduction in vital capacity, for example, may not produce speech abnormality. On the other hand severe or rapidly developing respiratory disorders may show a higher proclivity towards speech interference. In addition several specific speech features of respiratory impairment show selective impingement upon the speech mechanism".

The measurement of vital capacity is also useful in determining phonation quotient and vocal velocity index, apart from providing information regarding the total volume of air available for phonation.

Jayaram (1975) and Nataraja (1986) have reported that there was no significant difference between males of the normal and the dysphonic groups but a significant difference was found between the females of normal and dysphonic groups. Thus the measurement of vital capacity would help in differentiating dysphonics from normals. This parameter will be more important when there is a respiratory disorder along with or other conditions leading to voice disorders.

Mean air-flow rate

The air flow is important in bringing about vocal fold vibrations. The regulation of the air flow is basically involuntary and highly automatic in ordinary speech, but the public speakers or singers learn to rely heavily on a partial control of breathing mechanism (Boone, 1983).

"The aerodynamic aspect of phonation is characterised by four parameters. Subglottal pressure, supraglottal pressure, glottal impedance and volume velocity of the airflow at the glottis. The value of these parameters vary during one vibratory cycle according to the opening and closing of the glottis. These rapid variations in the values of the aerodynamic parameters cannot usually be measured in living human beings because of technical difficulties. For clinical purposes, the mean value of these parameters is usually determined" (Hirano, 1981).

The relationship between these parameters is shown as $P_{sub} - P_{sup} = MFR \times GR$, when P_{sub} is the mean subglottal pressure; P is the mean supraglottal pressure; MFR, the mean flow rate represented as unit of volume velocity; and GR, the mean glottal resistance. "Strictly speaking, the 'mean' used here implies the root mean square (rms) value" (Hirano, 1981).

The mean air flow rate is defined as the volume of air flow per unit of time, i.e.

$$\text{MFR} = \frac{\text{Total volume of air flowing during phonation}}{\text{Duration of phonation during which volume of air flow was measured}}$$

(Hirano, 1981)

Schneider and Baken (1984) have reported the influence of lung volume on the relative contributions of glottal resistance and expiratory force to the regulation of subglottal pressure. That is, lung volume influences the consistency and strength of the relationship between airflow, intensity and frequency. As noted earlier, the mean air flow rate is related to subglottal air pressure and glottal resistance and thus is related to frequency of vocal fold vibration and the intensity of voice. Hence the MFR varies with the pitch and register used (Brodnitz, 1959; Large, Iwata and Von Leden, 1972 and McGlone, 1967). In other words variations are found in MFR with variations in subglottal air pressure and glottal resistance.

Markel (1973) reported that the changes in intensity were accompanied by a proper balance between the force of subglottal air and the tension of glottic muscles. But there was direct relationship between the quantity of air passing through the larynx and increased vocal intensity.

Hirano, Miyahara, Hirose, Kiritani and Fujiyama (1970) reported the increase in MFR with increase in intensity. MFR varied with the frequency and intensity of the voice (Isshiki, 1959; Isshiki and Von Leden, 1964; Yanagihara and Koike,1967). Kunze (1964) and Isshiki (1964) have reported a flowrate of 100 cc/sec for normal phonation in the modal register. Yanagihara and Von Leden (1967) have reported ranges of 100 to 180 cc/sec in normal males.

Hirano (1981) presents the normal values of MFR of adults as reported by several investigations and he states that "the average values of MFR range from 81 to 141 ml/sec ... no consistant difference in MFR has been observed between the males and females either during maximum sustained phonation and the phonation over a comfortable period, or between results obtained either with the spirometer or pneumatograph. In most reports, the value ranges approximately from 70 to 200 ml/sec. The critical region, which indicates the possible ranges for the normal population is approximately 40 to 200 ml/sec. It appears reasonable to regard MFR values greater than 200 ml/sec or less than 40 ml/sec as abnormal, as far as phonation at a habitual pitch and loudness is concerned".

The MFR in different pathological conditions or dysphonics have also been studied.

Yanagihara (1969) has based on an analysis of data obtained from more than 100 cases, made the following diagnostic implications.

a. Flow rate more than 300 cc/sec with phonation time less than 50% suggests that low glottal resistance is the dominant contributing factor for the vocal dysfunction which may be diagnosed as hypofunctional voice disorder.

b. Flow rate upto 250 cc/sec with phonation time ratio of more than 70% and high phonation volume, vital capacity ratio suggests that a high glottal resistance is the dominant contributing factor for the vocal dysfunction which can be labelled as a hyper functional voice disorder. He stresses that the aerodynamic examination of phonation can be a valuable adjunct to other physiological studies for an understanding of laryngeal disorder.

Iwata, Vanladen and Williams (1972) found that the higher MFR correspond to hypotension conditions in the larynx, for example, in the unilateral paralysis of vocal cords, and lower MFR values were suggestive of hypertensive conditions such as contact ulcer, etc. Irregularities of the airflow during phonation were considered to be reflected in acoustic signals. These functions may be closely related to the pathological changes in the vocal cords, even in patients with apparently normal MFR. This suggests that the MFR during phonation and especially the degree of air flow fluctuations provide useful quantitative measures of laryngeal dysfunction (Iwata et al., 1972).

Iwata, Esaki, Iwami and Tasaku (1976) studied 73 dysphonics before and after surgery making use of aerodynamic and acoustic measurements. They concluded that the objective aerodynamic and acoustic examinations before and after surgery were useful in evaluating laryngomicro surgery. Hippel Meowinski (1978) studied the degree of air flow speed during phonation, the volume of air flow during phonation and the duration of phonation in 22 subjects with normal vocal function and 33 with voice disorders. In more than 50% of these cases the MFR has been found to be within the normal range. Regarding the MFR values in the cases of vocal nodules, polyps and polypoid swelling (Reinke's edema) of the vocal cords have been reported (Shigemori, 1977; Hirano, 1975; Saito, 1977). "In many cases the value of the MFR exceeds the normal range, but not as marked as in the cases with recurrent laryngeal nerve paralysis" (Hirano, 1981).

Shigemori (1977) reported a positive relationship between the MFR and the size of the lesion. A reduction in the MFR value has been found after surgical treatment of the lesion. According to Hirano (1981), the MFR values in the case of tumours of the vocal fold(s) most of which were neoplastic, varied from patient to patient. Isshiki and Von Leden (1964) report that the MFR always exceeded the normal range in the case of a large number.

As reported by Hirano (1981) the MFR values were within the normal range in the case of spastic dysphonia and contact granuloma. Jayaram (1975) reported a significant difference between the dysphonic and normal males, whereas female groups did not show any significant difference. Thus the studies have shown a relationship between vocal function and air flow measurements. They also have indicated that vocal function, normal or abnormal, can be assessed by the air flow measurements". The measurement of the mean air flow rate is often done as an outpatient procedure" (Hirano, 1981).

The following conclusions have been drawn by the investigators:

a. The parameters of the study were dependent upon the intensity of phonation.

b. The values of the normal group and the dysphonic patients with complete closure of the glottis during phonation were approximately the same and

c. The values of the dysphonic patients with incomplete closure of the glottis were significantly different from the values of the other two groups.

The MFR in recurrent laryngeal nerve paralysis, has been reported to be more than in normals. Such reports have been made by several investigators. A tendency of greater MFR has been reported with greater lateral fixation of the vocal cords. MFR has been considered to be a good indicator of the phonatory function in recurrent laryngeal nerve paralysis. And it has also been reported that MFR can be used as a monitor of treatment (Hirano et al., 1968; Hirano, 1975; Isshiki, 1977; Saito, 1977; and Shigemori, 1977).

Shigemori (1977) has reported MFR in 26 cases of sulcus vocalis as varying from 50 to 723 ml/sc. She has also reported that it was greater than 250 ml/sec in 12 patients (46%) and greater than 300 ml/sec in 6 patients (23%). MFR values in the cases of laryngitis have been reported by Isshiki and Vonleden (1964); Hirano et al. (1968); Iwata et al. (1972, 1976) and Shigemori (1977).

Phonation Quotient

An indicator of the vocal function is the ratio of vital capacity to maximum phonation duration (Sawashima, 1966). Hirano, Koike and Von Leden (1968) termed this ratio as phonation quotient. Hirano (1981) has defined this as "the value obtained when the vital capacity is divided by the maximum phonation time".

$$\text{Phonation quotient} = \frac{\text{Vital capacity (VL)}}{\text{Maximum Phonation Time (MPT)}}$$

A high correlational relationship between MFR and PQ in normal subjects has been reported by Hirano et al. (1965). The PQ has been considered an indicator of air usage and can be used when MFR cannot be directly determined as recommended by Iwata and Von Leden (1970).

The total volume of air used during maximum sustained phonation (PV) is usually less than the vital capacity (Yanagihara et al., 1966; Yanagihara and Koike, 1967; Isshiki, Okamura and Morimoto, 1967; Yoshioka, Sawashima, Hirose, Ushijima and Honda, 1977. Therefore it has been concluded by Hirano (1981) that PQ is usually larger than MFR during maximum sustained phonation.

The positive relationship between MFR measured during maximum sustained phonation and PQ has been confirmed by Iwata and Van Lden (1970) in pathological cases and by Yoshioka et al. (1977) in normal and pathological cases. However, a significant relationship has not been reported between PQ and MAF at comfortable duration of phonation (Shigemori, 1977; Yoshika et al., 1977).

Krishnamurthy (1986) reported a mean PQ of 131.16 in males and 123.23 in females. The PQ values have ranged from 78.75 to 187.37 cc/sec in males and 83.33 cc/sec to 183.33 cc/sec in females. The average PQ in normal adults has been found to be between 120 and 190 ml/sec. The upper limit of the normal range varied from 200 to 300 ml/sec between different reports. Most of the investigators reported on increased PQ value in most cases of vocal cord paralysis and abnormally high PQ have been found in cases of lesions of the vocal folds, including nodules, polyps, polypoid swelling and neoplasms (Hirano et al., 1968; Iwata and Vonleden, 1970; Shigemori, 1977 and Yoshioka et al., 1977). PQ has been considered to be useful in evaluating surgical treatment with certain cases of vocal pathology (Shigemori, 1977).

Maximum phonation duration

Maximum phonation duration or MPD has been suggested as a clinical tool for evaluation of vocal function for the past three decades. "A good criterion for the general quality of voice is immediately available by determining the phonation time" (Arnold, 1955). Gould (1975) said that maximum phonation duration measures give an indication of the overall status of laryngeal functioning and tension in the larynx and any neuromuscular disability. A short

1. Total air capacity available for voice production
2. The expiratory power and
3. The adjustment of the larynx for efficient air usage/ i.e. the glottal resistance.

The results of the study by Isshiki et al. (1967) indicated that none of the experimental subjects utilized the total vital capacity for phonation. The amount of air volume expired during longest phonation ranged from 68.7 to 94.5% of the subjects vital capacity. Yanagihara and Koike (1967) obtained similar findings with percentage ranging from 50 to 80% for males and from 45 to 70% for females. Lewis et al. (1982) found a significant and dominant relationship between vital capacity and the length of phonation of /a/. They also suggested that with twenty trials, the maximum phonation obtained would reflect utilization of a higher percentage of the vital capacity.

The amount and kind of training an individual had has been considered as yet another variable affecting the duration. Lass and Michael (1969) indicated that athletes generally did better than non-athletes and also trained singers were better than non-singers. However, the results obtained by Sheela (1974) showed no significant relationship between phonation duration of trained and untrained singers. The phonation duration ranged from 15 to 24 secs in trained phonation duration with a large air escape suggests a neuromuscular deficit such as laryngeal nerve paralysis.

'Norms' for maximum phonation duration vary from 10 sees for consonants in children to 30 sees for vowels in adults (Arnold, 1955). According to Van Riper and Irwin (1958) normal individuals should sustain a vowel for atleast 15 secs without difficulty. Fairbanks (1960) reported a duration of 20-25 secs as normal. The normal values for MPD have been reported by several investigators. The average is greater for males (25-35 sees) than for females (15-25 sees). Bless and Saxman (1970) studied MPD in boys and girls aged 8 and 9 years and found the MPD for girls was 19 sees and for boys it was 16 secs. These results were contrary to most of the other studies in that the girls had longer MPD than the boys. Further, the results obtained by Coombs (1976) in her study of children with varying degrees of hoarseness indicated no significant relationship between SQ and phonation duration. The difference may reflect the compounding aspects of hoarseness on the duration.

Shigemori (1971) investigated MPD in school children. The MPD was found to increase with age. The difference between males and females was not significant except among seventh grade children. Launer (1971) measured MPD for /a/, /u/ and /i/ in children aged 9 through 17 years. There was no statistically significant difference between the three vowels. Phonation duration increased with increasing age and boys had a longer sustained phonation time than girls. Lewis, Casteel and McMohan (1982) found no statistically significant relationship between phonation time and age using subjects of 8 and 10 years. However, Ptacek, Sander, Naloney and Jackson (1966) found that MPD decreased as a function of increasing age.

This lack of agreement among the results of different studies made several investigators to study variables which affect MPD. Variables investigated include vital capacity and air flow rate (Yanagihara et al., 1966; Brackett, 1971), vocal pitch and intensity (Ptacek and Sander, 1963; Yanagihara et al., 1966; Yanagihara and Koike, 1967), sex (Ptacek and Sander, 1963; Yanagihara et al., 1966; Yanagihara and Koike, 1967; Yanagihara and Vonleden, 1967; Coombs, 1976), age (Launer, 1971; Coombs, 1976) and height and weight (Launer, 1971).

Yanagihara and Koike (1967) indicated that the air volume available for maximally sustained phonation (i.e. phonation volume) varied in proportion to vital capacity and this was specific to sex, height, age and weight of individuals. They concluded that maximum sustained phonation was achieved by three physiological factors. They were:

singers and in untrained singers it ranged from 10 to 29 sec. Sawashima (1966) found no significant difference in the MPD in standing or sitting positions.

Sanders (1963) found MPD with twelve trials and found no difference between 1st and the 12th trial. Stare (1977) indicated that adults demonstrated greater maximum duration of /a/ when fifteen trials were used. Lewis et al. (1982) have found that the practice of utilizing three trials to determine the MPD was inadequate. They report that it was not until the 14t trial that 50% of their subjects produced the maximum phonation and not until the 17th trial, did all their subjects produce maximum phonation duration. The authors believed that this finding to be not only statistically significant, but also, more importantly clinically significant. However, most of the other studies have been based on three trials (Yanagihara et al., 1966; Yangihara and Koike, 1967; Yanagihara and Von Leden, 1967; Launer, 1971; Coombs, 1976).

Although many researchers have suggested the effect of height and weight to the length of phonation duration Luchsinger and Arnold (1965); Michael and Wendahl (1971), Lewis et al. (1982) found no statistically significant relationship between them.

Ptacek and Sander (1963) appear to be the first to suggest that the maximum duration of phonation may be influened by the frequency and sound pressure level of voice, then the male subjects could sustain phonation longer than females, especially at low frequencies and sound pressure levels. As both frequency and sound pressure level increased, the phonation duration between males and females tended to become more similar. However a considerable degree of variability among subjects was still evident in that significant differences existed for frequencies and sound pressure levels for male phonations, but not for female phonations. Inversely, the frequency sound pressure level interaction was significant for females but not for the males. Different results were found by Lass and Michael (1969). They report that for low frequency phonations of both males and females, and for the moderate frequency phonations of males, there was a general tendency for phonation duration to increase as a function of sound pressure level. However, in high frequency phonations for both males and females, there was a tendency for phonation duration to decrease as sound pressure level increased.

Yanagihara et al. (1966) and Yanagihara and Koike (1967) measured the maximum phonation duration at three different pitches - low, medium and high in normal adults.

Phonation duration was reduced at high frequencies for both males and females. The MPDs for males were 28.4 sec for low pitch, 30.2 sec for medium pitch and 23.7 sec for high pitch, while those for females were 21.7 sec for low pitch, 22.5 sec for medium pitch and 16.7 sec for high pitch.

Komiyama, Buma and Watanabe (1973) measured the maximum phonation duration taking amount of the intensity of the voice. Measurements were made at different pitches. The results indicated that the 'phonation duration' in a higher frequency range showed a lower value compared with the value of a lower frequency range. They also observed that the 'phonation capacity' by the integration of the voice intensity with phonation duration and reported that 'phonation capacity' diminished and showed a remarkable decrease in high frequency phonation during the register transition.

Maximum duration of phonation has been used as a diagnostic tool. A significant reduction below normal levels can be related to inadequate voice production. Arnold (1955) reports that in the cases of paralytic dysphonia, the phonation duration was always 3-7 secs. Hirano (1981) opined that clinically the maximum phonation time values smaller than ten sec should be considered abnormal.

Shigemori (1977) also reported that in pathological cases, abnormal findings were most evident in a measure of the MPD, than in the mean air flow (MAF) or phonation quotient (PQ) . An abnormally short MPD was found in cases of recurrent laryngeal nerve paralysis. The MPD varied depending on the cord position in laryngeal nerve paralysis (Shigemori, 1977).

Jayaram (1975) and Nataraja (1986) have reported a significantly lower MPD in a dysphonic group than in a matched normal group. Further while a significant difference in MPD was found between males and females in the normal group, no such difference was seen in the dysphonic group. These results are similar to those reported by Coombs (1976) where no significant difference was observed with respect to MPD, between males and females with hoarseness.

Ptaeek and Sander (1963) appear to be the first to relate the MPD to the perception of breathiness. Although none of the voices of their subjects were considered non-normal, they were able to divide their subjects into two groups - long phonators and short phonators. When these two groups were judged as to degree of breathiness from least to most on a seven point scale they found that the long phonators tended to be judged as having less breathiness than the short phonators. In addition perceived breathiness decreased as a function of increase in intensity, and high frequency phonations tended to be rated as more breathy, tan corresponding low frequency phonations.

Von Leden, Yanagihara and Werner (196) showed that short MPD was associated with laryngeal pathology and can be improved by treatment. They reported an increase in phonation duration from 1.33 to 14.79 sec in one case and 3.91 to 8.66 secs in another case (both of whom had unilateral vocal fold paralysis after injecting teflon into the affected folds). Michael, Kircner Shelton and Hollinger (1968) also demonstrated an increase in the phonation duration from 4 sec to more than 20 sec as a result of teflon treatment of unilateral vocal fold paralysis.

Shigemori (1977) reported that MPD is valuable for monitoring the effects of surgical treatment in selected disorders of the larynx, especially in recurrent laryngeal nerve paralysis, sulcus vocalis, nodules and polyps.

Arnold (1955) has stated that MPD as a measure demonstrates the general status of the patients respiratory coordination, but more accurately indicates the relative efficiency of the pneumolaryngeal interaction.

Recently Indian studies carried out by Krishnamurthy and Jotinder (1994), Salaj (1994), Rajeev (1995) have shown a significant difference between normal males and females in the maximum phonation duration.

Fundamental frequency in phonation

Fundamental frequency is the lowest frequency that occurs in the spectrum of a complex tone. In voice also, the fundamental frequency is considered the lowest frequency in voice spectrum. This keeps varying depending on several factors. This is considered to be determined by the frequency of vocal fold vibration.

"... both quality and loudness of voice are mainly dependent upon the frequency of vibration. Hence it seems apparent that frequency is an important parameter of voice" (Anderson, 1961). There are various objective methods to measure the fundamental frequency of the vocal cords/voice.

The study of fundamental frequency obviously has important clinical implications. Cooper (1974) used spectrographic analysis, as a clinical tool to determine and compare the fundamental frequency in dysphonics before and after vocal rehabilitation. Jayaram (1975) and Shantha (1973) found a significant difference in habitual measure between normals and dysphonics.

A study conducted by Asthana (1977) to find the effect of fundamental frequency and intensity variation on the degree of nasality in cleft-palate speakers showed that the cleft palate speakers had significantly less nasality at higher frequency levels than at the habitual frequency. But the degree of perceived nasality did not change significantly when the frequency was lowered. Thus it is apparent that the measurement of the fundamental frequency is important in the diagnosis and the treatment of voice disorders.

Fundamental frequency in speech

An evaluation of the fundamental frequency in phonation may not represent the habitually used fundamental frequency of the individual. Many investigators have studied fundamental frequency as a function of age and in various pathological conditions. Different types of speech samples, i.e. phonation, reading, spontaneous speech and singing have been used in recent studies. In literature one often finds comparison of results of different studies. But it is not clear whether the same type of speech samples have been considered for such comparisons. And further it is not clear whether all the speech samples would yield the same results. However, clinical experience has shown that the subjects use different fundamental frequencies, under different conditions. To verify this clinical impression an experiment was conducted by Nataraja and Jagadeesh (1984). They measured fundamental frequency in phonation, reading, speaking and singing and also the optimum frequency in thirty normal males and thirty normal females. They observed that the fundamental frequency increased from phonation to singing with speaking and reading in between.

Table 1 shows the results of the study

Group		Optimum frequency	Fundamental frequency in			
			Phonation	Speaking	Reading	Singing
Male	M	127.91	141.98	166.00	192.25	211.17
	SD	8.48	29.72	33.84	43.27	60.64
Female	M	241.13	237.03	266.26	272.91	304.04
	SD	18.39	29.30	53.72	44.98	55.89

Thus the results of the above study cautions that fundamental frequency has to be measured under different conditions in the evaluation of voice disorders, i.e. it may not be enough if one considers one conditions to determine the means fundamental frequency used by the case for evaluation of voice.

The age dependent variations of mean SFF reported by Bohme and Hecker (1970) indicate that the mean SFF decreases with age upto the end of adolescence in men. In advanced age, the mean SFF becomes higher in men but slightly decreases in women. Hudson and Holbrook (1981) invested the mean modal frequency, in reading, in two hundred young black adults age ranged from 18 to 29 years and found it to be 110.15 Hz in males and 193.10 Hz in females compared to a similar white population studied by Fitch and Holbrook (1970) the black population had a lower mean modal frequency.

No significant differences in the mean and median SFF between laryngitic and non-laryngitic voices has been reported by Shipp and Huntington (1965). Murray (1978) studying the SFF characteristics of four groups of subjects, namely vocal fold paralysis, benign mass lesion, cancer of the larynx and normals, noted that the parameters of mean SFF failed to separate the normals from the three groups of pathologic subjects. In a study, Murray and Dohesty (1980) reported that along with other voice production measures such as directional and magnitudinal perturbation the SFF improved the discriminant function between normal voice and voice of patients with malignancy of the larynx. Sawashima (1968) reported a rise in the mean SFF in cases of sulcus vocalis and a fall in mean SFF in cases of polypoid vocal folds and virilism. Very high mean SFF values resulted from disturbances of mutation in males. According to Hirano (1981) "at present, MSF (the SFF) is measured as a clinical test value".

Frequency range in phonation and speech

Hudson and Halbrook (1981) studied the FF range in reading in a group of young black adults, age ranging from 18 to 29 years. A mean range from 81.95 to 158.80 Hz in males and from 139.05 to 266.10 Hz in females was found. Compared to a similar white population studied by Fitch and Halbrook (1970), it was found that the black population had greater mean frequency ranges. Fitch and Halbrooks (1970) white subjects showed a greater range below the mean than above. The behaviour was reversed for the black subjects.

McGlove and Hollicn (1963) report that in women 65 to 79 years the speaking pitch variability changes little. However, Stoicheff (1981) reported an increase in variability of FF in post-menopausal adults, which was interpreted as indicating decreased laryngeal control over FF adjustments.

General conclusions about the diagnostic value of FF variability are difficult to make because such measurements are helpful in certain pathological conditions but not in others (Kurt, 1976). Shipp and Huntington (1965) indicated that laryngitic voices had significantly smaller ranges than did post-laryngitic voices. Murray (1978) found reduced semitone ranges of SFF in patients with vocal fold paralysis, as compared with normals. Murray and Dohesty (1980) concluded from another study that the variability in SFF, along the directional and magnitudinal perturbation factors, enhanced the ability to discriminate between speakers with no known vocal pathology and speakers with cancer of the larynx.

Fluctuations in fundamental frequency and intensity in sustained phonations

Presence of small perturbations or irregularities of glottal vibration in normal voice has long been known (Moor and Vonleden, 1958; Vonleden, Moore and Timcke, 1960). Relatively few attempts have been made to noted the purturbations in F_Q and intensity, although such measures may have value in describing the stability of laryngeal control (Liberman, 1963).

Aperiodic larygneal vibrations have been related to abnormal vocal production by various investigators (Carhart, 1938, 1941; Bowler, 1964). Vanleden et al. (1960) report that in a frame by frame analysis of ultra-high speed motion pictures "the commonest observation in pathological conditions is a strong tendency for frequent and rapid changes in the regularity of vibratory pattern".

Lieberman (1963) found that pitch perturbations in normal voice never exceeded 0.5 msec in magnitude in the steady state portion of long sustained vowel. Similar variations in the fundamental periodicity of the acoustic waveform have been measured by Fairbanks (1970). The results reported by Lieberman (1963) were confirmed by Iwata and Vonleden (1970) and the 95% confidence limits of pitch perturbations in normal subjects ranged from -0.19 to 0.2 msec. "The cycle to cycle variation in period that occurs when an individual is attempting to sustain phonation at a constant freqency" has been termed as 'jitter'.

There are a number of methods for obtaining jitter measurements and when actual measurements are made, a number of alternative methods of data reduction are available to the investigators (Heiberger and Horii, 1982). While considering the neurophysiological significance of jitter, Heiberger and Horii (1982) state that "physiological interpretations of jitter in sustained phonation should probably include both physical, structural variations and myoneurological variations during phonation. A number of high speed laryngoscopic motion pictures revealed the laryngeal structures (vocal folds) are not totally symmetric. Different amounts of mucus accumulates on the surface of the folds during vibration. In addition, turbulent air flow at the glottic also cause some perturbations limitations of laryngeal servo mechanism through the articular myotic and mucousal reflex systems (Gould and Okamura, 1974; Wyke, 1967) may also introduce small perturbations in laryngeal muscle tones. Given without the consideration of the reflex mechanisms, the laryngeal muscle tones have inherent perturbation due to the time-staggered activities of motor units that exist in many voluntary muscle contractions (Baer, 1980).

Wilcox (1978), Wilcox and Horii (1980) report that a greater magnitude of jitter occurs with advancing age and this they attribute to the reduced sensory contributions from the laryngeal mechanoeceptors. However, these changes in voice with age may be due to physical changes with respiratory and articulatory mechanisms. Heibeyer and Horii (1982) while considering the perceptual significance of jitter state that even though these acoustic measures have been viewed as some of the physical correlates of 'rough' voice quality there is discrepancy between "the findings of the earlier synthesis studies (Coleman, 1969; Coleman and Wendahl, 1967; Wendahl, 1963, 1966a, 1966b) and the more recent human voice studies (Horii, 1979; Ludlow et al., 1979). The synthesis studies found near perfect correlations between jitter and perceived roughness. The human voice studies on the other hand showed low, non-significant correlations between magnitude of jitter and the perceived roughness levels.

In pathological speech, Lieberman (1963) indicated that the magnitude of the perturbation factor (magnitude of pitch perturbation exceeding 0.5 msec) might be useful in detection of laryngeal diseases. Smith and Liberman (1964) also investigated the relation between pitch perturbations and pathological conditions of the larynx. Koike (1973) studied pitch perturbation and amplitude modulation in different types of vocal attacks and emphasised that these studies might give more detailed information about pathological changes in the larynx. Iwata (1972) tested the voice of 20 normal subjects and 27 patients with various laryngeal diseases for pitch perturbations. The results showed that correlograms were useful in differentiating normal and abnormal voice and different types within the abnormal group.

Shimmer, in any given voice is dependent atleast upon the modal frequency level, the total frequency range and the SPL relative to each individual voice (Michael and Wendahl, 1971). Ramig (1980) postulated that Shimmer and jitter values should increase when subjects are asked to phonate at a specific intensity and/or as long as possible. Whitehead and Emanuel (1974) found vocal fry productions were perceived to be relatively rough compared to the modal register phonations and manifested elevated spectral noise levels, comparable to those associated with simulated abnormally rough phonations. This was explained by Wendahl (1963, 1966) and Coleman (1969) who indicated that when two audible complex waves manifest equal amounts of wave aperiodicity, the wave with the higher fundamental frequency will tend to be heard as least rough.

Jitter and Shimmer have been applied to the early detection of laryngeal pathology. Liberman (1961, 1963) states that the pitch perturbation factor might be a useful index in detecting a number of laryngeal diseases. Koike (1969) showed that a relatively slow period modulation of vowel amplitude was observed in patients with laryngeal neoplasms. He reasoned from this that the measurement and analysis of such modulation might be useful in assessing laryngeal pathology. Crystal and Jackson (1970) measured both the fundamental frequency and amplitude perturbations of voice in persons with varying laryngeal conditions and concluded that several purely statistical measures of the data they extracted might be useful as guidelines in detecting laryngeal dysfunction. Koike (1973) investigated the pitch periods of voice produced by pathologic speaker, and found that discrimination between laryngeal tremor and paralysis was possible. The perturbation factors, during sustained vowels, were significant in discriminating normal talkers from those with laryngeal cancer (Murry and Doherty, 1980).

Vonleden and Koike (1970) found a significant correlation between subjects with various laryngeal diseases and different types of amplitude modulations and affirmed the potential values of short term perturbations in the acoustic signal for diagnostic purposes. The data suggested and different types of amplitude modulations which in turn correlated with clinical groupings.

Kitagima and Gould (1976) studied the vocal shimmer during sustained phonations in normal subjects and patients with laryngeal polyps and found the value of vocal shimmer to range from 0.04 dB to 0.21 dB in normals and from 0.08 dB to 3.23 dB in the case of vocal polyps. Although some overlap between the two groups was observed they noted that the measured value may be a useful index in screening for laryngeal disorders or for diagnosis of such disorders and differentiation between the two groups was observed they noted that the measured value may be a useful index in screening for laryngeal disorders or for diagnosis of such disorders and differentiation between the two groups.

Shipp and Huntingson (1965) recorded the voice of 15 subjects while each had acute laryngitis, and when their voice returned to normal. The recordings of laryngitic and post-laryngitic voices were subjected to a number of perceptual evaluations and to fundamental frequency measurements. The results indicated that the laryngitic condition received higher mean hoarseness ratings than did the normal condition. Laryngitic voices had significantly smaller ranges of frequency than did the post-laryngitic voice. A small number of frequency breaks were also observed in laryngitic voice.

The fluctuations in frequency and intensity in a given phonation sample may indicate the physiological (neuromuscular) or pathological changes in the vocal mechanism. Kim et al. (1982) have analyzed the vowel /e/, (earlier analyzed by Imaizumi et al., 1980) using the spectrograph, in 10 voices of patients with recurrent laryngeal nerve paralysis and 10 normals to obtain the following acoustic parameters.

The acoustic parameters obtained from the spectrographs were:

1. Extent of fundamental frequency fluctuation: The extent of fluctuation was defined as the per cent score of the ratio of the peak to peak value of fluctuation (F_0) to the mean fundamental frequency (F_0).

2. Speed of fundamental frequency fluctuation: This has been defined as the number of positive peaks within 1 sec.

3. Extent of amplitude fluctuation: This has been defined as the peak to peak value in decibel measured on an average amplitude display.

4. Speed of amplitude fluctuation, which was defined as the number of positive peaks on an amplitude display within 1 sec. Peaks of 3 dB or greater from adjacent troughs have been counted.

The results of this study have indicated that among the acoustic parameters studied, significant differences were found between the control and the diseased groups in terms fluctuation of F_0.

Yoon et al. (1984) studied the voice of patients with glottic carcinomas, using the same procedure and the parameters as described by Kim et al. (1982). They concluded that significant difference were found between the normals and patients with advanced carcinomas in terms of extent of fluctuation and speed of amplitude fluctuations.

Rashmi (1985) concluded that

1. Fluctuations in frequency of the initial and final segments of phonation of /a/, /i/ and /u/ showed a decreasing trend with age in males.

2. The 14 to 15 year old group showed an increase in the range of fluctuations for all the vowels.

3. In females, there was a decrease in the range of fluctuations in frequency of the initial and final segments in upto the age of 9 years, an increase in the range of fluctuations in the 9-11 years old females, which again drops down till the age of 15 years.

4. The medial segment of phonation, both males and females was quite steady, and the range of fluctuations as a function of age did not show much difference.

5. No difference in the ranges of fluctuations in frequency between males and females was observed in the younger age groups.

6. The males consistently showed greater fluctuations in frequency in the phonation of /a/, /i/ and /u/ than the females of 14-15 year old group.

7. The fluctuations in the initial and final segments of phonation for all the three vowels was greater than the fluctuations in the medial segment, for both males and females.

8. The fluctuations in intensity did not show any systematic trend for any vowels both in males and females. However, the initial segment of phonation showed a significantly layer fluctuation in intensity in the above 12 year groups, in the cases of males, for all three vowels.

Higgins and Saxman (1989) reported higher value of frequency perturbation in males than females. Gender differences nay exist not only in the magnitude but also in the variability of frequency perturbation. Linville and Korabic (1987) have found that intra-speaker variability tend to be greater on the low vowel /a/, with less variability on high vowels /i/ and /u/.

Several investigators have studied the measure of amplitude perturbations in normal and pathological groups. Vanaja (1985), Tharmar (1991) and Suresh (1991) have reported that as age increased there was an increase in fluctuations in frequency and intensity in phonation and these differences were more marked in females. Nataraja (1986) found that speed of fluctuation in fundamental frequency and extent of fluctuation in intensity parameters were sufficient to differentiate dysphonics from normals.

Huggins and Saxman (1989) investigated within subject variation of three vocal frequency perturbation indices over multiple sessions for 15 female and 5 male young adults (pitch perturbation quotient and directional perturbation factor). Coefficient of variation for pitch perturbation quotient and directional perturbation factor were considered indicative cf temporal stability of these measures. While jitter factor and pitch perturbation quotient varied considerably within the individual sessions, while directional perturbation factor was a more temporally stable measure.

Venkatesh et al. (1992) reported the jitter ratio (JR), relative average perturbation 3 point (RAP3), deviation from linear trend (DLT), shimmer in dB (SHIM) and amplitude perturbation quotient (APQ) to be most effective parameters in differentiating between normal males, normal females and dysphonic groups. They also said that pitch perturbation (3 point) and DLT are most useful in differentiating laryngeal disorders. Chandrashekhar (1987) found no significant difference in jitter values in /a/ for males and /i/ and /u/ for females when compared to dysphonics. Also the shimmer values were greater for vocal module cases than in normals with respect to both male and female groups. But the values were significant only for males. On the whole, he found significant differences in jitter and shimmer values between normals and dysphonics.

Vieira, Mclnnes and Jack (1997) assessed the agreement between jitter obtained through acoustic and EGG measurements in pathologic voices, acoustic and EGG waveforms of sustained vowels (/i/, /a/, /u/) produced by 15 dysphonic patients. The agreement between acoustic and EGG derived jitter was poorer for /i/ and /u/ than for /a/

vowels. For /a/ vowels a method of acoustic jitter estimation was proposed that combines peaks and zero crossings and resulted in increased consistency.

Hence the present study was undertaken to

1. to verify the usefulness of the acoustic and aerodynamic parameters put forth by Nataraja (1986) in differentiating various groups of pathological voices.

2. To device an objective method for classification of different voice disorders.

Artificial Neural Networks (ANNs) commonly referred to as Neural Networks (NNs) are also termed as neuro-computers, connectionist networks, parallel distributed processors, etc. In the most general form, a neural network is a machine designed to model, the way in which the brain performs a particular task or function of interest by using electronic components or simulated in software on a digital computer.

ANN is a massively parallel distributed processor that has a natural propensity for storing experimental knowledge and making it available for use. It resembles the brain in two respects: (1) Knowledge is acquired by the network through a learning process, (2) Interneuron connection strength, known as weights are used to store knowledge. In other words, AMNs are "biologically" inspired networks having the apparent ability to initiate the brains activity to make decisions and draw conclusions when presented with complex and noisy information (Haykins, 1994).

Characteristics of Neural Networks (NNs)

NN derives its computing power through its massively parallel distributed structure and its ability to learn and therefore to generalize. Generalization refers to the NN producing reasonable outputs for input not encountered during training (learning). These two information processing capacities make it possible for NNs to solve complex (large scale) problems that are currently intractable. A complex problem of interest is decomposed into a number of relatively simple tasks and NNs are assigned subsets of tasks (eg. pattern recognition, associative memory, control) that match their inherent capabilities (Haykin, 1995).

The use of NNs offers the following useful properties and capabilities (Zurada, 1992).

1. **Non- linearity:** A neuron is basically a nonlinear device. Consequently, a NN, made up of interconnections of neurons, is itself nonlinear and is distributed throughout the network. Nonlinearity is a highly important property,

particularly if the underlying physical mechanism responsible for the generation of an input signal (eg. speech signal) is inherently nonlinear.

2. Input-output mapping: A popular paradigm of learning called supervised learning involves the modification of the synaptic weights of a NN by applying a set of labelled training samples or task samples. Each example consists of a unique input signal and the corresponding desired response. The network is presented an example picked at random from the set, and the synaptic weights (free parameters) of the network are modified so as to minimize the differences between the desired response and the actual response of the network produced by the signal in accordance with an appropriate statistical criterion. The training of the network is repeated for many examples in the set until the network reaches ad steady state, where there are no significant changes in the synaptic weights. The previously applied training examples may be reapplied during the training session but in a different order. Thus, the network learns from the examples by constructing an input-output mapping for the problem at hand like the supervised learning paradigm which suggests a close analogy between the input-output mapping performed by a NN and non-parametric statistical inference.

3. Adaptivity: NNs have built in capacity to adopt their synaptic weights to changes in the surrounding environment particularly, a NN trained to operate in a specific environment can be easily retrained to deal with minor changes in the operating environmental conditions. When operating in a non-stationary environment (i.e. one whose statistics change with time), a NN can be designed to change its synaptic weights in rule time. The natural architecture of a NN for pattern classification, signal processing and control applications coupled with adaptive capability make it ideal tool for adaptive pattern classification, adaptive signal processing and adaptive control.

4. Evedential response: In the context of pattern classification, a NN can be designed to provide information not only about which particular pattern to select, but also about the confidence in the decision made. Thus information may be used to reject ambiguous patterns if arised and thereby improve classification performance of the network.

5. Contextual information: Knowledge is represented by the very structure and activation state of the NN. Every neuron in the network is potentially affected by the global activity of all other neurons in the network. Consequently, contextual information is dealt with naturally by a NN.

6. Fault tolerance: A NN implemented in hardware form, has the potential to be inherently tolerant in the sense that its performance is degraded under adverse operating conditions (Bolt, 1992) . That is, if a neuron or its connection links are damaged, recall of stored pattern is impaired in quality. However owing to the distributed nature of the information in the network, the damage has to be extensive before the overall response of the network is degraded seriously.

7. VLSI implementability: The massively parallel nature of a NN makes it potentially fast for the computation of certain tasks. This feature makes it ideally suited for implementation using very-large scale integrated (VLSI) technology. This provides a means of capturing truly complex behaviour in a highly hierarchical fashion, which makes it possible to use a NN as a tool for real time applications involving pattern recognition signal processing and control.

8. Uniformity of analysis and design: Basically, NNs enjoy universality as information processors. That is, the same notation is used in all the domains involving the application of NNs, in terms of sharing theories and learning algorithms, etc.

9. Neurobiological analogy: The design of a NN is motivated by analogy with the brain, which is a living proof that fault tolerant parallel processing is not only physically possible but also fast and powerful. Neurobiologists took to ANNs as a research tool for the interpretation of neurobiological phenomena. For eg: NNs have been used to provide insight on the development of promotor circuits in the oculo-motor system (responsible for eye movements) and the manner in which they process the signals (Robinson, 1992).

Structure and function of ANN

The typical structure of an ANN is given in figure 1 below. ANNs are basically three layer system consisting of:

1. input layer of nodes for data entering the network.
2. hidden layer between the input and output layers.
3. output layer of nodes that produce the networks output responses.

Typical structure of ANN

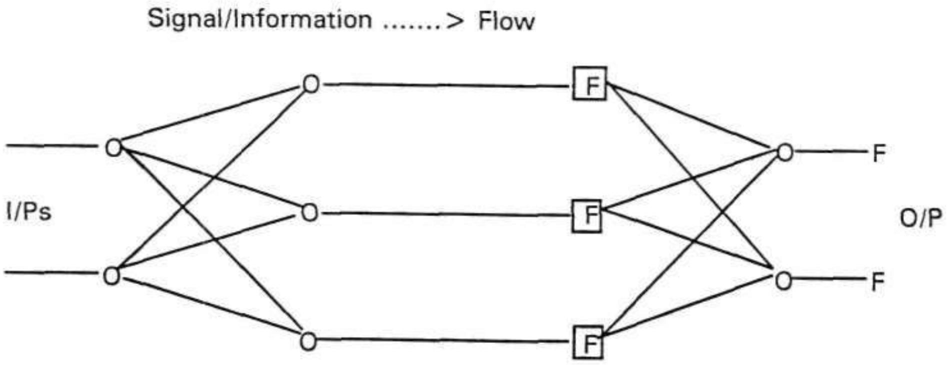


FIGURE 1

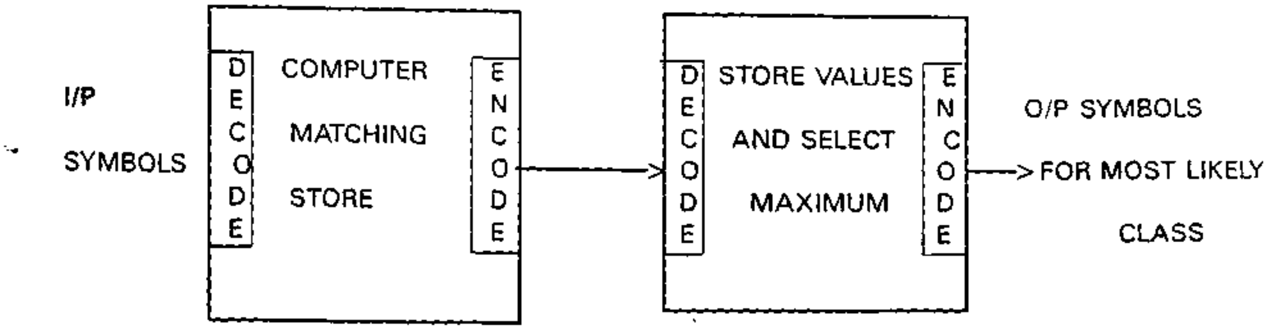
McCullock and Pitts, regarded as pioneers in the field of NNs outlined the first model of an elementary neuron in 1943 (Roal and Mankame, 1996). However they could not realise the model using the bulky vacuum tubes of the era. A learning algorithm scheme for updating a neurons connections (weights) was proposed by Donald Hebb inl949. A new and powerful learning law called Widrow-Hoff learning rule was developed by Bernard Widrow and Mavrian Hoff in 1960.

The ANN structure is trained with known samples of data, the known pattern or information in the form of digital signals. Then the ANN is ready to recognise a similar pattern when it is presented to the network. The nonlinear characteristics of ANN are due to the non-linear activation function 'F'. This is useful for accurate modelling of non-linear systems. That is, in the non-linear systems the output variables do not depend on the input variables in a linear manner. The dynamic characteristics of the system itself would depend on either one or more of the amplitude of the input signal, its waveform or its frequency.

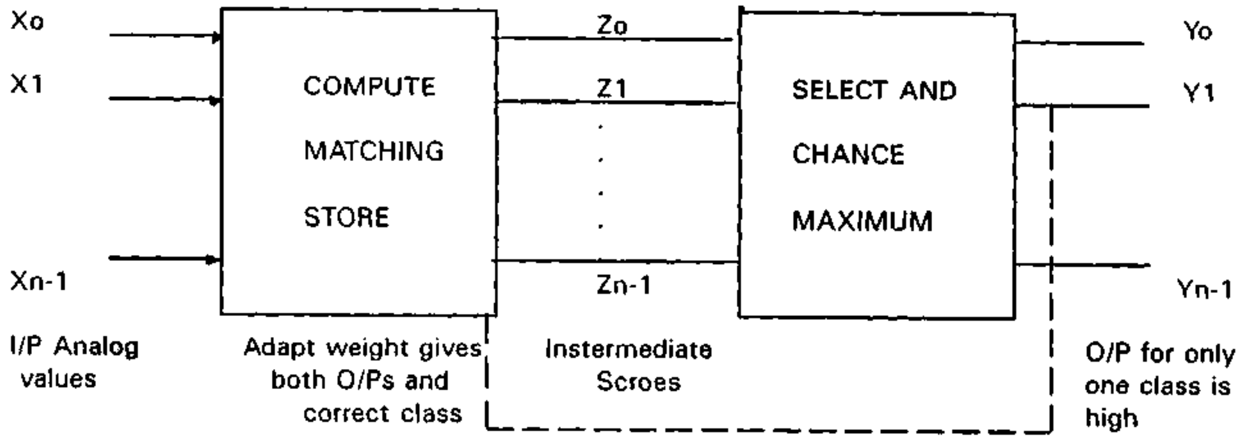
The structure of an ANN is a set of processing units (nodes) arranged in rows. Inputs nodes are interconnected by simple connections with an internal layer of hidden nodes and a single output mode. Rather than having a fixed algorithmic approach to a classification problems, an ANN is sequentially presented with a set of supervised training cases input data passes with correct output. The ANN modified its behaviour in this process of training by adjusting the strengths or weights of the connections until its own output conveys to the known correct output. The information "learned" by the ANN is stored in the weight of the network gives to the connections between nodes. Thus ANNs are designed to realise very specific computational tasks/ problems by the highly inter-connected, parallel computational structures with many and relatively simple elements.

Neural net and traditional classifiers

Block diagrams of traditional and neural net classifiers are given in the figure as shown.



TRADITIONAL CLASSIFIER FIGURE 2



NUERAL NET CLASSIFIER

FIGURE 3

Both traditional and NN classifiers determine which of M classes is most representative of an unknown static input pattern containing N input elements. In a speech recogniser the input might be the output envelope values from a filter bank spectral analyzer sampled at one time instant and the classes might represent different vowels.

Inputs and outputs of the traditional classifier are passed serially and internal computations are performed sequentially. Parameters are typically estimated from training data nd then held constant.

In the adaptive NN classifier input values are fed in parallel to the first stage via N input connections. Each connection carries an analog value which may take on two levels for binary inputs or may vary over a large range for continuous valued inputs. The first stage computes matching scores and outputs these scores in parallel to the next stage over M analog output lines. Here the maximum of these values is selected and enhanced. The second stage has one output for each M classes. After classification is complete, only that output corresponding to the most likely values will be on strongly or 'high'; other inputs will be 'low'. In this design outputs exist for every class and this multiplicity of outputs must be preserved for further processing stages as long as the classes are considered distinct. In the simplest classification system these output lines might go directly to lights with labels that specify class identities. In most complicated cases they may go to further stages of processing where inputs from other modalities or temporal dependencies are taken into consideration of the correct class is provided, then this information and classifier outputs can be fed back to the first stage of the classifier to adapt weights using a learning algorithm as shown in figure 3. Adaptation will make a correct response more likely for succeeding input patterns that are similar to the current pattern.

The parallel inputs required by the neural net classifiers suggest that real time hardware implementations should include special purpose pre-processors. One strategy for designing such processors is to build physiologically-based pre-processors modelled after human sensory systems. Pre-processor filter banks for speech recognition that are crude analogs of the cochlea have been constructed (Martin, 1970). More recent physiologically based processor algorithms for speech recognition attempt to provide information similar to that available on the auditory nerve (Delgutte, 1984). Many of these algorithms include filter bank spectral analysis, automatic gain control, and processing which uses timing or synchrony information in addition to information from smoothed filtered output envelope.

Classifiers shown in the figures 2 and 3 perform three different tasks:

1. They can identify which class best represents an input pattern, where it is assumed that inputs have been corrupted by noise or some other process. This is a classical decision theory problems.

2. They can be used as a content addressable or associative memory, where the class exempler is designed and the input pattern is used to determine which exempler to produce. This is useful when only part of the information is available and the complete pattern is required as in bibliographic retrieval of journal references from practical information.

3. They can perform vector quantize (Makhoul et al., 1985) or cluster the N inputs into M clusters. Vector quantizers one used in image and speech transmission systems to reduce the number of bits to transmit the analog data. In speech and image recognition applications they are used to compress the amount of data that must be processed without loosing important information on either application, the number of clusters can be pre-specified or may be allowed to

grow upto a limit determined by the number of nodes available in the first stage.

Multi-Layer Perceptron

These are 'feed forward' nets with one or more layers of nodes between the input and output nodes. These additional layers contain hidden units or nodes that are not directly connected to both input or output nodes. They overcome many of the limitations of single layer perceptions and are more effective with the recent development of new training algorithms and have been successful for many problems of interest.

The capabilities of MLPS stem from the non-linearities used within the nodes. It has been demonstrated that no more than three layers are required in perception-like 'feed-forward' nets because three layer net can generate arbitrarily complex decision regions (Lippman, 1987). The number of nodes in the second layer must be greater than one when decision regions are disconnected or meshed and cannot be formed from one convex area. The number of second layer nodes required in the worst case is equal to the number of disconnected regions in input distributions. The number of nodes in the first layer must typically be sufficient to provide three or more edges for each convex area generated by every second layer node. There should thus be more than three times as many nodes in the second as in the first layer. This is centered primarily on multilayer perceptions with multiple output nodes when sigmoidal non-linearities are used and the decision rule is to select the class corresponding to the output node with the largest output. The behaviour of these nets is more complex because decision regions are typically bounded by smooth curves instead of by straight line segments and analysis is thus more difficult.

Uses of ANNs

Some of the uses of ANNs (Raol and Mankame, 1996) include:

- Information storage/recall: The recall is a process of decoding the previously stored information by the network.
- Pattern recognition/classification: To recognise a specific pattern from a cluster of data/to classify sets of data or information.
- Non-linear mapping between high dimensional spaces (mathematical modelling of non-linear behaviour of systems).
- Time-series prediction (like weather forecasting), modelling of non-linear aerodynamic phenomena, detection of faults/failures in systems like power plants and aircraft sensors.

Metz, Schiavetti and Knight (1992) used an Artificial Neural in estimating the speech intelligibility of hearing impaired subjects from known acoustic variables.

Leinonen, Kangas, Torkkola and Juvas (1992) examined the vowel [a:] in a test word (finnish) which was judged as normal or dysphonic by two speech language pathologists, with the self-organizing map, the artificial neural network algorithm of Kohenen. The algorithm produced a two-dimensional representation (maps) of speech. Input to the acoustic maps consisted of 15-component spectral vectors calculated at 9.83 msec intervals from short-time power spectra. They found that the dysphonic voices deviated from normals both in composition of the short-time power spectra (characterised by the dislocation of the trajectory pattern on the map) and in the stability of the spectrum during the performance (characterised by the pattern on trajectory on the map). They concluded that this method is suitable not only for diagnosis but also for therapeutic purposes.

Mujunen, Leinonen, Kangas and Torkkola (1993a) studied word initial samples of fricative [s] preceding vowels [a:], [ae], [e:], [i:], [u:], [o:] and [y:] in finnish words with a self organizing map, the neural network algorithm of Koheenen. The [s] samples were drawn from 10 women, aged 20-45 years. The subjects were selected on the basis of having normal [s] articulation. Fifteen component input vectors, which constituted the input to the acoustic map was calculated from short-time FFT (Frequency Fourier Transform) spectra at 9.83 msec intervals. In all the 10 subjects the [s] samples preceding the round vowels [u:] and [o:] clearly differed from the samples in front of unrounded [a:], [ae:], [e:] and [i:]. There were no significant differences between the locations of [s] proceeding [a], [u] and [y], or between those preceding [a], [e], [i] and [æ].

Mujunen, Leinonen, Kangas and Torkkola (1993b) further studied whether the self organizing map also distinguishes between misarticulated and normal [s]. They collected speech samples from 11 women aged 16-38 years, who had misarticulation for the [s] sound as examined by a speech therapist. A psychoacoustic evaluation was also done where the recorded speech samples of the subjects were classified as acceptable or unacceptable by 2 speech pathologists. The results showed that the map patterns of the most deviant [s] samples differed from those of the normal ones and those judged acceptable by most listeners. The degree of audible acceptability correlated well with the location of the sample on the map. They concluded that the self organized maps are suitable for the extraction and measurement of acoustic features underlying psychoacoustic classifications, and for on-line visual imaging of speech.

Farley (1994) created an artificial neural network model which was supposed to correspond to one component of fundamental frequency (F_0) control by the brain. Good F_0 control could be achieved using only seven neurodes. These included three motor nodes, a single inhibitory neurode influencing only the thyroarytenoid (TA) motor neurode, and three exitatory neurodes, one of which excited all motor neurodes and two of which excited the inhibitory neurode. The potential utility of this type of model in the study of mechanisms of vocal control was discussed.

Leinonen, Hihunen, Laakso, Rihkanen and Poppius (1997) obtained a perceptual reference for acoustic feature selection. 94 male and 124 female voices were categorised using the ratings of 6 clinicians on visual analog scales for pathology, roughness, breathiness, strain, asthenia and pitch. Partial correlations showed that breathiness and roughness were the main determinants of pathology. The 6-dimensional ratings (the 6 median scores for each voice) were categorized with the aid of the Sammon map and the self-organizing map. The 5 categories created differed with respect to the breathiness or roughness ratio and the degree of pathology.

Sunil, Murty, Venkatesh and Vijaya (1995) devised a study in the Indian population which attempted to evaluate the automatic recognition of electroglottographic (EGG) patterns for differential diagnosis of 128 laryngeal disorders using Artificial Neural Network. A comparison was also made between identification abilities of automatic pattern recognition methods, trained speech language pathologists, and student clinicians. Results indicate that automatic recognition of EGG patterns is possible with a recognition rate of 63.3%. The recognition rates of the trained speech pathologists and student clinicians were found to be poorer.

Hence from the above review of literature it is evident that different acoustic and aerodynamic parameters could be helpful in the differential diagnosis of various voice disorders. Also it can be seen from the review of Artificial Neural Networks (ANNs) that these software systems are very useful in building up models and they are mainly helpful in modelling complex, noisy and varying da

We know that acoustic and aerodynamic measures of voice show a wide variety of variation in normals and in pahologocal populations and also depend on the age and sex of the speaker.
Hence the present study aims to create a model to differentially diagnose voice disorders using an Artificial Neural Network. The neural network used in th present study

is Multi Layer Perception (MLP). Initially the study aims at training the neural network system, by using acoustic and aerodynamic parameters as the input and the diagnostic labels of various voice disorders as the output. Once the network is trained and a model is created, the efficiency of the model is to be checked, where the input consisting of acoustic and aerodynamic parameters are given and the network model has to predict the output or to which diagnostic category does the voice disorder belong.

METHODOLOGY

The purpose of the present study is to classify different voice disorders in terms of their acoustic and aerodynamic parameters using Artificial Neural Networks.

The study was carried out in two phases.

a. Collection of data on acoustic and aerodynamic parameters in normal and dysphonics.

b. Validation of data by comparing the data with other ANN for classification of voice disorders based on aerodynamic and acoustic parameters.

Phase I

The following acoustic and aerodynamic parameters were considered for different voice disorders as input to the neural network.

Aerodynamic parameters

1. Mean air flow rate
2. Maximum duration of phonation

Acoustic parameters

1. Fundamental frequency in phonation
2. Extent of fluctuations in fundamental frequency
3. Speed of fluctuations in fundamental frequency
4. Extent of fluctuations in intensity

5. Speed of fluctuations in intensity
6. Frequency range in phonation
7. Intensity range in phonation
8. Fundamental frequency range in speech.

SUBJECTS

The data was collected from 30 normals and 414 dysphonics.

The normal subjects were taken who had no history of a speech, language or hearing disorders. A total of 30 normal subjects were taken (N = 30). The normal group consisted of 15 males in the age range of 20 to 25 years and 15 females who were in the age range of 20 to 25 years.

The dysphonics subjects were those who visited All India Institute of Speech and Hearing with a complaint of voice problems. A total of 444 dysphonics were taken. The group consisted of 160 males in the age range of 5-75 years and 55 females in the age range of 11-65.

Those subjects who had been diagnosed as cases of 'voice disorder' after the routine otolaryngological, speech, psychological evaluation and audiological evaluation were included as subjects of this group.

The distribution of cases in terms of their diagnosis are shown in the table (Table 1).

Diagnosis	Males	Females	Total
1. Mass on the vocal folds (includes vocal nodules, vocal polyps and thickening of vocal folds)	114	47	163
2. Glottic chink (includes Glottic and chink and paralysis)	102	67	169
3. Congestion	50	22	72

Table 1: Table showing the distribution of different types of dysphonics both males and females

The study was devised in two stages. In the first stage the various acoustic and aerodynamic parameters were extracted and in the second stage these parameters were fed into the neural network for training and prediction.

Phase I: Procedures used to measure different parameters

I. Aerodynamic parameters

1. Mean air flow rate (MAFR)

Mean airflow rate has been defined as the amount of air collected in one second during phonation at a given frequency and intensity. In other words,

2. Measurement of maximum duration of phonation (MPD)

Maximum duration of phonation has been defined as the duration for which an individual can sustain phonation. This measurement was carried out in one of the sound treated rooms of the speech laboratory.

The subject was instructed as follows:

"Take a deep breath and then say /a/ as long as you can, with the voice that you usually use. Please try to maintain it at a constant level". The procedure was demonstrated. Then the subject phonated as long as possible. Using a stop watch, the duration of phonation of /a/ was measured. The subject was asked to repeat the whole process twice with 2-3 minutes gap between the trials. Thus for each subject 3 trials were given. The one which had the longest duration of the three trials were considered the maximum duration of phonation for that subject.

Acoustic parameters

All the acoustic parameters were computed in the following manner.

Instrumentation

The following instruments were used.

1. Dynamic microphone (AHUJA AUD-5354)
2. Pre-amplifier (PHILIPS PRE AMP-60)

Diagnosis	Males	Females	Total
1. Mass on the vocal folds (includes vocal nodules, vocal polyps and thickening of vocal folds)	114	47	163
2. Glottic chink (includes Glottic and chink and paralysis)	102	67	169
3. Congestion	50	22	72

Table 1: Table showing the distribution of different tvnes of dysphonics both males and females

The study was devised in two stages. in the first stage the various acoustic and aerodynamic parameters were extracted and in the second stage these parameters were fed into the neural network for training and prediction.

Phase I: Procedures used to measure different parameters

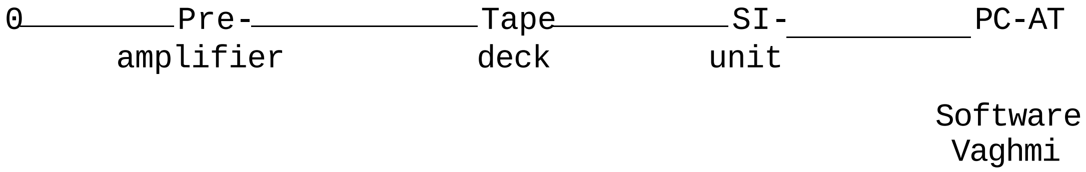
I. Aerodynamic parameters

1. Mean air flow rate (MAFR)

Mean airflow rate has been defined as the amount of air collected in one second during phonation at a given frequency and intensity. In other words,

3. Sony tape deck (TCFX 170)
4. Speech interface unit
5. PC-AT (486 DX) Vaghmi Software (Voice and Speech Systems, Bangalore)

Block diagram of the instrumentation setup



Speech sample

Sustained phonation of three vowels /a/, /i/ and /u/ were used in order to measure mean, fundamental frequency, frequency range, intensity range, extent and speed of fluctuations in frequency, extent and speed of fluctuations in intensity.

Recording of speech sample

The subjects were seated comfortably in the sound treated room. The dynamic microphone was kept in front of the subject at a distance of about 15 cm from the mouth. They were instructed to take a deep breath and say /a/. They had to maintain a constant intensity and pitch at a comfortable level. The output of the mic was fed to a Sony stereocassette deck with Hi-Fi CrO_2 cassette for recording the speech samples.

The speech samples were recorded at a recording speed of 17/8 ips, similar records were carried out for /i/ and /u/.

Analysis

The tape recorded speech signal was played back to the input of the SI unit.

The sample was digitized and fed through the interface unit at a sampling rate of 16 kHz using the programme 'Record' of VSS software. Before digitizing, the each sample was passed thruogh the alaising filter at 3.5 kHz with role off of 48 dB per octave. Digitized data was stored on the hard disk of the computer (HCL PC with Pentium 200 MHz processing). The level indicator of the speech interface unit was used to monitor the intensity level of the signal to avoid any distortion while digitizing the signal.

"VSS-Vaghmi" program (Inton) was used to extract F_0 and related measures.

Statistical analysis

A Mann-Whitney 'U' test was used to statistically analyze the data and to compare the subgroups of normal males and females and dysphonic males and females, with each other and between the groups.

PHASE II

Phase II of the study was carried out in the "Neural connection" software. The software requires the following basic requirements.

—-> An IBM compatible PC with a pentium 200 MHz processor.

--> Microsoft Windows 95

—-> (16 MB) RAM

--> A hard disk with 2 GB space

--> A SVGA monitor, with appropriate graphics card

—-> A mouse

The graphic user interface in the software provided the workspace, the program window where problem solving applications could be built. The right side of the screen in the workspace gave the tool modules. The tool modules included tools for data input and output, data manipulation statistical analysis, classification and modelling and forecasting.

The neural connection consisted of the following tool menus:

--> **Data input** - Imports data either from files or by cutting and pasting from other windows applications can be used to edit data.

--> **Outputting tools**

-—> Data output - exports data to files, displays results as text, and shows the success rate.

—-> Text output - Displays results as text and shows the success rate.

--> Graphics output - Displays a graph showing how the output from an application varies when two of the inputs vary.

—-> What if - Displays a plot of how the output varies when two of the inputs vary and gives a description of how the change in one input affects the other.

--> Time series plot - Displays the results of time series prediction against time.

Tools for modelling and forecasting

--> Multi layer perceptron - A neural network modelling tool that was optimized for prediction applications.

—-> Radial Basis function - A neural network tool that can be optimized for prediction and classification applications.

--> Bayesian network - A neural network modelling tool that can be optimized for prediction and classification applications.

--> Kohenen network - A neural network modelling that can be optimized for clustering applications.

—-> Closest class means classifier - A statistical modelling tool that could be used for classification applications.

--> Regression - A statistical modelling tool that could be used for prediction problems.

— > Principal component analysis - A statistical modelling tool that could be used to reduced complexity of an application.

Tools for filtering the data

--> Filter - A tool that allows to filter the data, apply mathematical functions, and examine the statistical distribution of the data.

--> Network combiner - A tool that allows to join two or more other tools together.

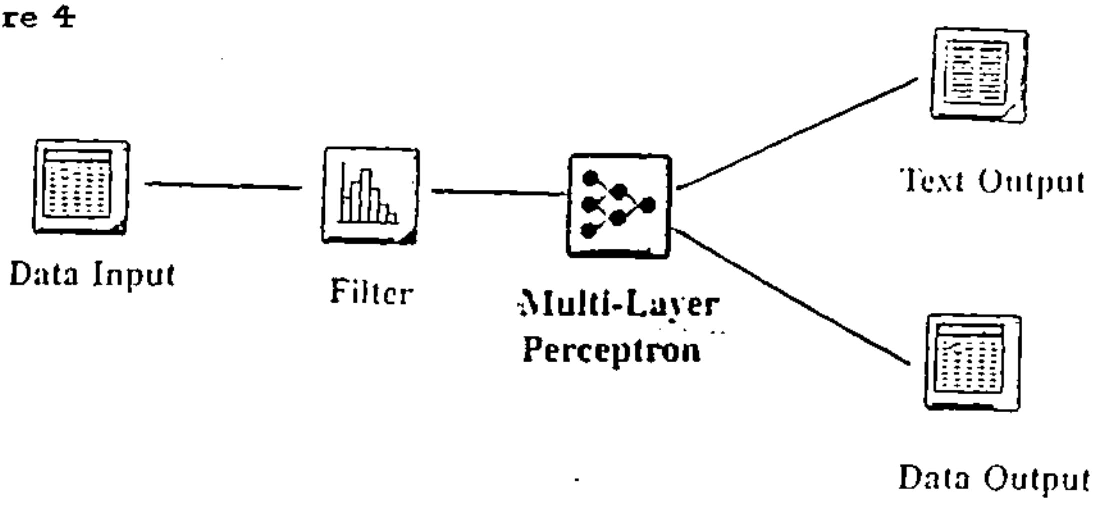
--> Simulator - A tool that produces the data format needed by the Graphics output and what if? tools.

--> Time series window - A tool that allows to manipulate time. series data in order to create windows before modelling takes place.

The neural network option chosen for the present study was '**Multi Layer Perceptron' (MLP)** along with the neural network the following tools were used in the following order:

1. Data input
2. Filter
3. Multi Layer Perceptron
4. Text Output
5. Data Output

Figure 4



The figure shows the order of arrangement of the tools which were used for the study.

All the tools could be moved into the work space by using the mouse. Initially the data input tool was brought to the work space and in order to input the training data into the tool the following steps were carried out:

1. Click the data input tool once. The tool menu appears.
2. From the data input tool choose "view". A blank spread sheet appears.
3. From the file menu choose "open". The data input dialog box appears. This box allows you to specify the developmental data file and run data file, which are loaded seperately. The developmental data file will be used to train and test the application. The run data file will be used when the model is implemented.
4. In the Development Data input group, click "Flat file".
5. Click "configure". A file open window appears.
6. The file name is typed. The present data is been stored in the spss data format (.sav). The training data used for the study was recorded in the file format "neural net. sav".
7. After selecting the file, click "OK".
8. After this the spread sheet appears.

The table (table 2) shows a portion of the spread sheet with the input and output (target) variables. In the spreadsheet

I	I	I	I	I	I	I	I	I	I	I	I	T
AGE	SEX	Fo	FoRs	PFo	PIo	EXTF	EXTI	SPEEF	SPEEI	MAFR	PHODU	PD
23	F	225	45.5	8.32	2.15	0.00	0.00	0.00	0.00	140	25	a
20	M	145	82.0	30.13	11.85	3.25	5.32	8.52	10.89	200	17	b
18	M	160	25.3	12.58	3.22	12.32	9.78	0.00	2.33	400	3	c
35	M	185	33.41	14.83	8.00	7.32	1.22	5.83	3.33	190	16	e
40	F	240	18.32	21.32	15.25	18.32	9.88	0.00	0.00	160	15	e
70	M	160	60.23	7.89	3.85	15.57	3.22	4.48	2.27	328	4	c
19	F	280	11.23	18.33	10.81	0.00	0.00	0.00	0.00	280	12	b
25	F	155	41.32	5.39	2.81	33.32	15.38	10.97	3.22	195	19	b

Table 2: **Table showing the Distribution of Input and Output Variables to be fed into the Neural Network**

Fo = Fundamental Frequency of Phonation
FoRs = Fundamental Frequency Range in Speech
PFo = Fundamental Frequency Range in Phonation
PIo = Intensity Range in Phonation
EXTF = Extent of Fluctuations in Frequency
EXTI = Extent of Fluctuations in Intenstiy
SPEEF = Speed of Fluctuations in Frequency
SPEEI = Speed of Fluctuations in Intensity
MAFR = Mean Airflow Rate
PHODU = Phonation Duration
PD = Provisional Diagnosis of Voice Disorder

a = Normal
b = Mass
c = Glottic Chink
e = Congestion

I = Input Variables
T = Target

each columns which contain different input variables are called fields. There are 12 input fields which includes age, sex and the acoustic and aerodynamic parameters. The 13 field is the target field which contains the diagnosis of different voice disorders.

For the sake of convinience the output data which consists of the diagnosis of various voice disorder had been given symbolic representations as given below:

a = Normal
b = Mass on the vocal folds (includes, vocal nodule, vocal polyps and thckening)
c = Glottic chink (includes Glottic chink and paralysis)
e = Congestion

When the data is entered inside the data input tool a major portion of it is allocated for training the data and a smaller percentage of the data is used for validating and testing the data. Hence we have

Training data - used to train the application.

Validation data - use to monitor the network performance during training.

Test data: Used to measure the performance of a trained application.

For training the limits are:

____> A maximum of 750 fields.

____> A maximum of 15,000 records.

Out of a total of 436 data used in the study 337 datas were used for training, 43 for validation and 45 for testing.

For running the model, the limits are:

____> A maximum of 750 fields.
____> A maximum of 32,000 records.

Out of a total of 436 data used in the study 337 datas were used for training, 43 for validation and 45 for testing.

When the spread sheet occurs different codes appear in the second coloumn from the left, they are:

T = Training data
X = Test data
V = Validation data
R = Run data

Also various codes appear on top rows of the screen:

I = Input fields
T = Target fields

R = Reference fields, they are passed through the application to be displayed on the output but not used or changed.
* = unused fields. There are not passed through the application.

The filter is important to preprocess the data when using a modelling and forecasting tool. Data fields can be switched on or off by the filter. When they are selected the filter passes hem to sucessor objects when requested. When they are not selected, the data are unavailable to successor objects.

The filter also helps in weighting the data. The data can be weighted in several ways to maximize their usefulness to the application. These weightings cannot be used with symbolic data, nor can logrithms or natural logrithms be used with integer data.

Also each field can for analysed by the filter. The filter shows a histogram and several statistical measures such as mean, median, mode, variance, standard deviation, skewness, kurtosis, skewness and quantiles. If a particular field is badly skewed appropriate weighting function can be applied in the filter to process the data.

The data for the study was analysed in the filter to see the characteristics of the input acoustic and aerodynamic variables. When the histograms for individual data fields were obtained it was found that the histograms for fundamental frequency range, intesity range, and speed of fluctuations in frequency and intensity were badly skewed. The histograms are show in figure.
The logrithmic function sqrt(x) was applied to fundamental frequency range and square root function sqrt(x) was applied to speed of fluctuations in intensity to make then distributions normal. The intensity range was badly skewed and so it was not passed through the application.

Multi layer Perception is the neural network model which is being used for the present study. To train the Multi layer Perceptron:

——> From the Multi layer Perceptron menu, choose train.

The multi layer perceptron (MLP) is trained by an incremental learning technique where by it is trained in stages. In the first stage a sample of examples from the training set is used to train the multi layer perceptron. The best network in this stage is then passed to the 2nd stage and is used as a starting point for training. In 2nd stage, a larger sample of data is used to train the network, and best network is passed on to the next stage. This procedure continues for four stages. The number of samples trained in each stages are:

Stage I	-	100
Stage II	-	100
Stage III	-	100
Stage IV	-	10,000

The training automatically stops when the MLP finds a global minimal error which is equivalent to 0.001 for the training and validation data or it stops when the number of sample in the 4th stage comes to 10,000.

Training can also be stopped by clicking 'STOP' in the MLP performance window.

To improves the results of the MLP performance the following changes can be carried out.

- ——> The number of hidden layers can be changed - presently MLP in the 'Neural connection' software has two hidden layers.

- ——> The number of nodes in each layer - the number of modes in the hidden layers can be changed. Increasing the number of nodes in the hidden layers beyond a certain extent decreases the ability of the neural network to generalize based on the model created.

- ——> The transfer or activation, function used by the nodes. The activation functions are of 3 types:
  - ——> Tanh function
  - ——> Sigmoid function
  - ——> Linear function

- ——> The learning algorithm used by the MLP these are of two types:
  - ——> Conjugate gradient
  - ——> Steepest descent

Data output tool Helps us to view the trained data and examine the results. The data sets are identified by coloured codes in the left hand coloumn on the screen.

T = Training data, in cyan

V = Validation data, in bright green

X = Test data, in yellow

R = Run data, in pale blue

To view the results in the data output tool, the data output window is opened select 'view' and from the view menu choose the training, validation, test or run data to be viewed.

The Text output tool helps the results to be viewed in a simple format that can be easily understood. It shows the results in percentage as to how well the neural network has performed during the training, validation and test phases.

After the model was implemented, the 'Run data file' was used to further check the performance of the neural network.

In order to access the data into the 'Run data file':

1. Click the data input tool once. The tool menu appears.
2. From the data input tool choose "view". A spread sheet appears.
3. From the file menu choose "open". The data input dialog box appears.
4. In the run data input group, click "Flat file".
5. Click "configure". A file window appears.
6. The file name is typed. The rundata was recorded in the format "ANN.sav".

7. After selecting the file, click "ok".

8. After this the spread sheet of the rundata appears.

The run data consisted of a total of 8 subjects, 2 each for the four groups taken for the study. In the run data only the aerodynamic and acoustic parameters were given without the diagnosis and the network model had to predict the diagnosis of the voice disorder.

In order to run the rundata into the application model created:

1. From the data output menu choose "view".
2. From the file name choose "setup".
3. Select "flat file" from the run data
4. Press the "configure" button.

After the rundata is set up in the data output tool:

1. From the data output menu, choose "view".
2. Select "Output Run Data". Neural connection asks to confirm whether the specific run data file has to be outputed.
3. Click "yes".

The results appear in the spread sheet in the last column named target "T".

RESULTS AND DISCUSSION

The purpose of the present study was to create a model using an artificial nueral network which could classify a normal and abnormal voice and also differentiate between the types of voice disorders.

The following acoustic and aerodynamic parameters were taken up which constituted the input data to the neural network.

Aerodynamic Parameters

1. Mean air flow rate (MAFR)
2. Maximum phonation duration (MPD)

Acoustic Parameters

1. Fundamental frequency for phonation
2. Extent of fluctuations in frequency
3. Speed of fluctuations in frequency
4. Extent of fluctuations in intensity
5. Speed of fluctuations in intensity
6. Fundamental frequency range in phonation
7. Intensity range in phonation
8. Fundamental frequency range in speech

Before these parameters could be fed in to the neural network a statistical analysis was done for each of these acoustic and aerodynamic parameters to find whether these parameters could significantly differentiate a normal and an abnormal voice.

The results of different parameters have been discussed, after analyzing them using appropriate statistical tests.

Aerodynamic parameters

1. Mean air flow rate: The examination of tables (3.1 and 3.2) and graph (1) revealed that there was a significant difference between normals and dysphonics, both in males and the females.

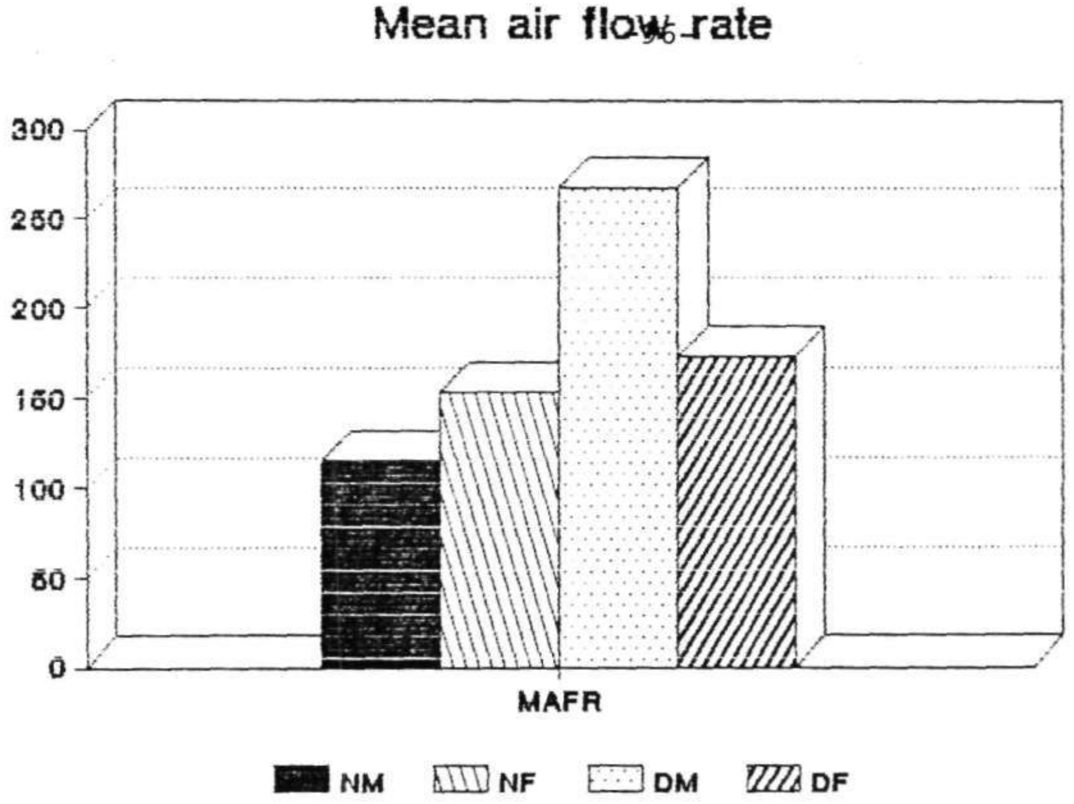
Groups	Mean	S.D.	Range
NM	114.4	21.36	92-133
DM	267.90	179.28	62.5-500
NF	153.5	30.96	90-230
DF	173.5	77.49	80-423

Table 3.1: The mean, S.D. and range of mean airflow rate, in normal and dysphonic groups, both males and females.

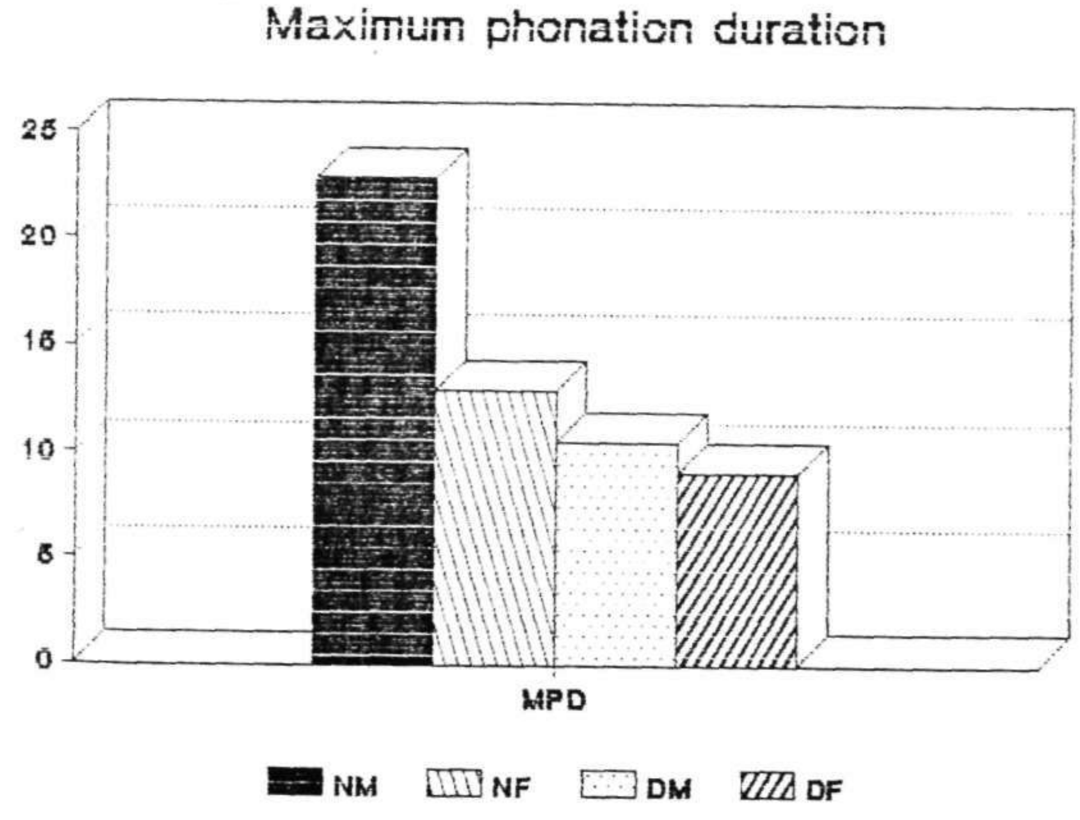
Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	+

(+ > significant of 0.05 level)

Table 3.2: Comparison of normal vs dysphonics, both males and females in terms of mean air flow rate.



GRAPH 1



GRAPH 2

NM - Normal Male; NF - Normal Female; DM - Dysphonic Male;
DF - Dysphonic Female

The dysphonic groups showed a much higher mean air flow rate and a much greater variability than the normals. The subjects of the dysphonic groups showed a maximum mean air flow rate of 500 cc and 375 cc which were higher than the maximum in normals. The dysphonic males showed a much higher mean air flow rate than the dysphonic females whereas no difference was seen in normals. Henc the hypothesis stating that there is no significant difference between normals and dysphonics both males and females for mean airflow rate is rejected and the hypothesis stating that there is no significant difference between

a. normal males and normal females

b. dysphonic males and dysphonic females

Similar findings have been reported by several investigators (Isshikki and Von leden, 1964; Hirano et al., 1968; Yoshioka et al, 1977; Shigemori, 1977; Jayaram, 1975; Nataraja, 1986). Hence the values obtained for this parameter in this study were similar to those obtained in other studies in normals and dysphonics. This parameter was considered useful in differentiation of dysphonics from normals.

2. Maximum phonation duration:

The study of tables (tables 4.1 and 4.2) and graph (Graph 2) showed that the maximum phonation duration was significantly lower in the dysphonic group than in the normal group. Both the dysphonic males and females had almost the same duration of maximum phonation, where as normal males showed a much longer phonation duration than normal females.

Groups	Mean	S.D.	Range
NM	22.8	5.35	18-29
DM	10.43	5.78	10-16
NF	12.87	2.22	3-25
DF	9.12	3.70	6-19

Table 4.1: The mean, S.D. and range of MPD in normal males and females and dysphonic males and females.

Group	Significance
NM vs NF	+
NM vs DM	+
DM vs DF	-
DF vs NF	+

(+ significant at 0.05 level)

Table 4.2: Comparison of normal vs dysphonics, both males and females in terms of maximum phonation duration.

Hirano et al; (1968), Jayaram (1975), Shigemori (1977) and Nataraja (1986) also report that shorter duration in dysphonics than normals in terms of phonation durations were observed in different types of voice disorders.

Statistical analysis showed significant difference between

a. normal males and females

b. dysphonic males and females

c. between dysphonic females and normal males.

But no significant difference was observed between males and females of dysphonic groups. Thus values of this parameter obtained in this study was similar to the previous studies.

Hence the hypothesis stating that there was no significant difference between normals and dysphonics both males and females and between normal males and females for maximum phonation duration is rejected and the hypothesis stating that there is no significant difference between dysphonic males and dysphonic females is accepted. This parameter has been found to be useful in differentiating dysphonics from normals.

Acoustic Parameters

3. Fundamental frequency in phonation: (Fo)

The fundamental frequency in phonation as per the tables (tables 5.1 and 5.2) and Graph (graph 3) was different for the normal and the dysphonic groups in both males and females. These differences were statistically significant. The males and females of the dysphonic group showed greater variations than the males and females of the normal group. Similar findings have been reported by Shantha (1973), Jayaram (1975), Cooper (1979) and Nataraja (1986).

Groups	Mean	S.D.	Range
NM	117.0	25.36	100-159
DM	152.15	48.73	208-263
NF	230.0	19.89	87-237
DF	200.96	42.9	150-285

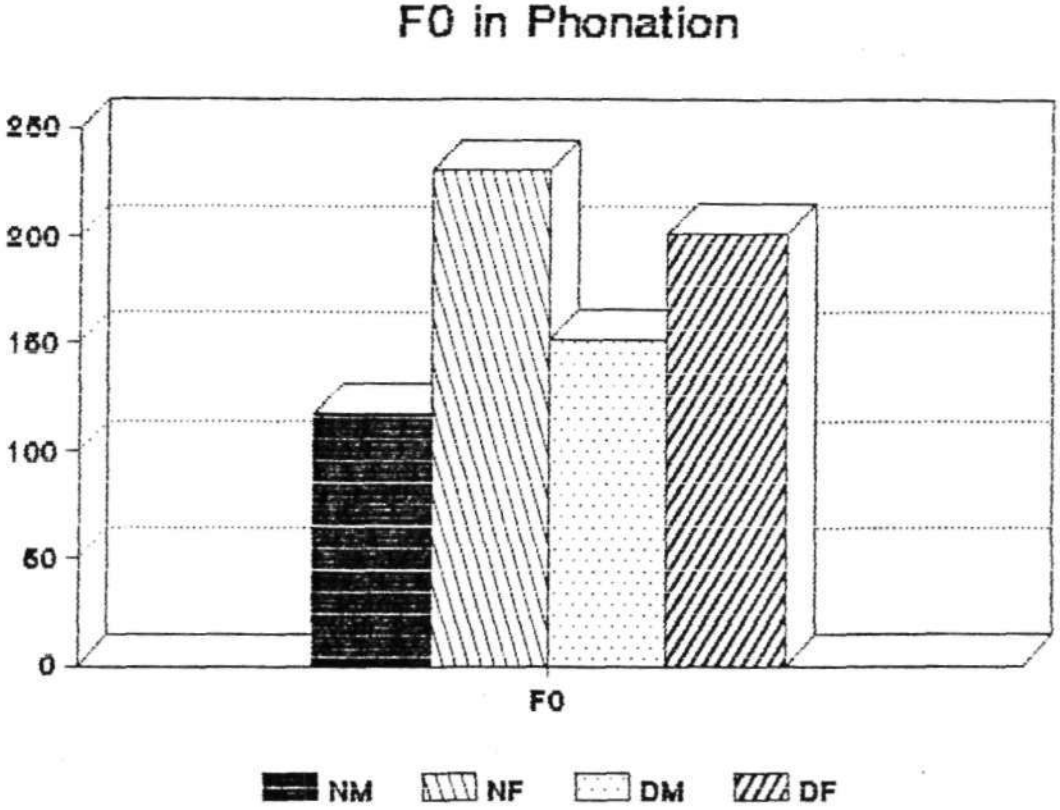
Table 5.1: The mean,S.D. and range of fundamental frequency in phonation in dysphonics and normals, males and females.

Group	Significance
NM vs NF	+
NM vs DM	+
DM vs DF	-
DF vs NF	+

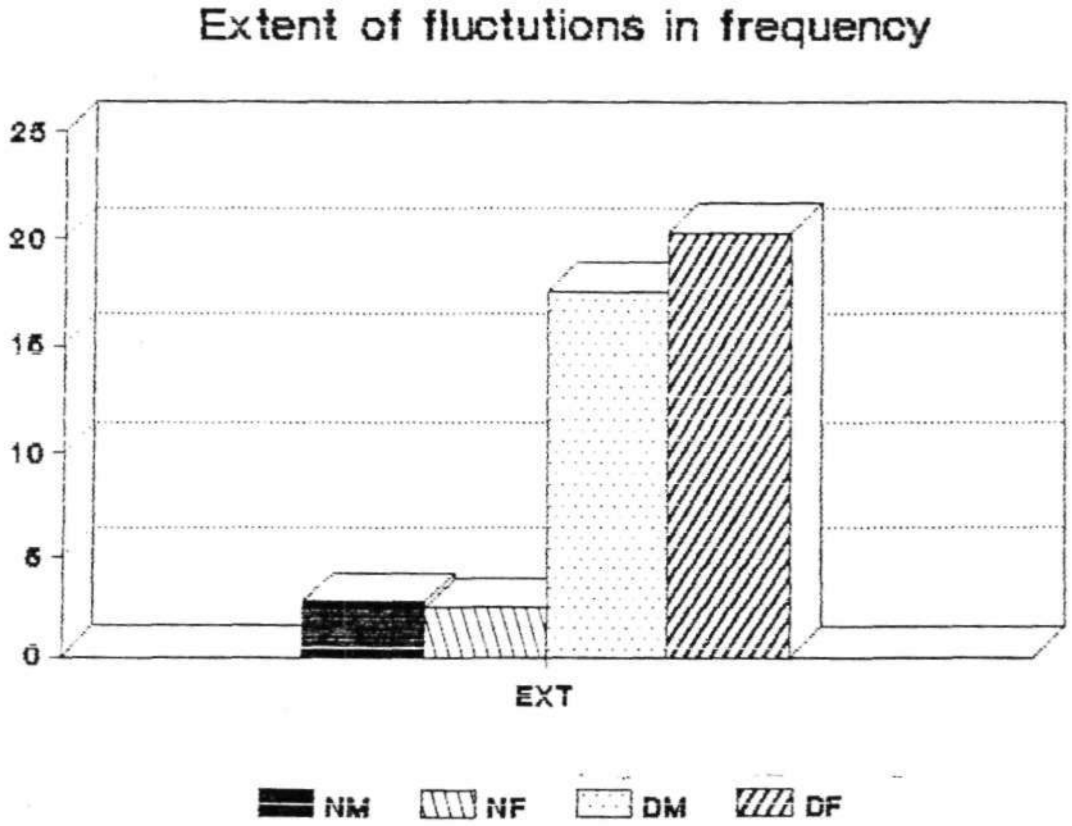
(+ significant at 0.05 level)

Table 5.2: Comparison of normal vs dysphonics, both males and females in terms of fundamental frequency in phonation.

Thus the hypothesis stating that there is no significant difference between normals and dysphonics both males and females, and between (a) normal males and females



GRAPH 3



GRAPH 4

NM - Normal Male; NF - Normal Female; DM - Dysphonic Male;
DF - Dysphonic Female

(b) dysphonic males and females for fundamental frequency in phonation is rejected.

Hence the values of this parameter obtained in this study was similar to those obtained in the previous studies. This parameter was found to be useful in differentiating normals from dysphonics.

4. Extent of fluctuations in frequency

The results for both the groups, the normal and the dysphonic, on extent of fluctuation in fundamental frequency in phonation are given in Tables (Tables 6.1 and 6.2) and Graph (Graph 4).

Study of these tables and graph indicated that the dysphonic males and females had higher values than normal males and females. Further variability was also found to be more in subjects of dysphonic group compared to normal group.

Groups	Mean	S.D.	Range
NM	2.74	2.5	0-5
DM	17.50	10.28	0.4-22.0
NF	2.41	0.74	3-81.9
DF	20.30	17.34	3.77-62.00

Table 6.1: The mean, S.D. and range extent fluctuation in fundamental frequency in phonation for both groups.

Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	-

(+ significant at 0.05 level)

Table 6.2: Comparison of normal vs dysphonics, both males and females in terms of extent of fluctuation in frequency in phonation.

The results and statistical analysis of this parameter showed that the two groups were significantly different from each other, both in males and females. The means and standard deviations of the males and the female dysphonics were much higher than those of the normal males and females. In both groups no differences were seen between the males and females.

Hence the hypothesis stating that there is no significant difference between normals and dysphonics both males and females for extent of fluctuations in frequency is partially accepted and the hypothesis stating that there is no significant difference between

a. normal males and normal females

b. dysphonic males and dysphonic females for extent of fluctuations for frequency is accepted.

The present results are in agreement with the studies of Kim et al; (1982), Yoon et al; (1984), Imaizumi et al; (1980) and Nataraja (1986). The values of the parameter obtained in this study were similar to various obtained in the previous study. Thus this parameters was found to be useful in distinguishing dysphonics from normals.

5. Speed of fluctuations in fundamental frequency

It was found that dysphonic males and females had a greater number of fluctuations than normal males and females. The results are shown in the tables (Tables 7.1 and 7.2) and Graph (Graph 5). The variations were much higher in the dysphonic group than normals.

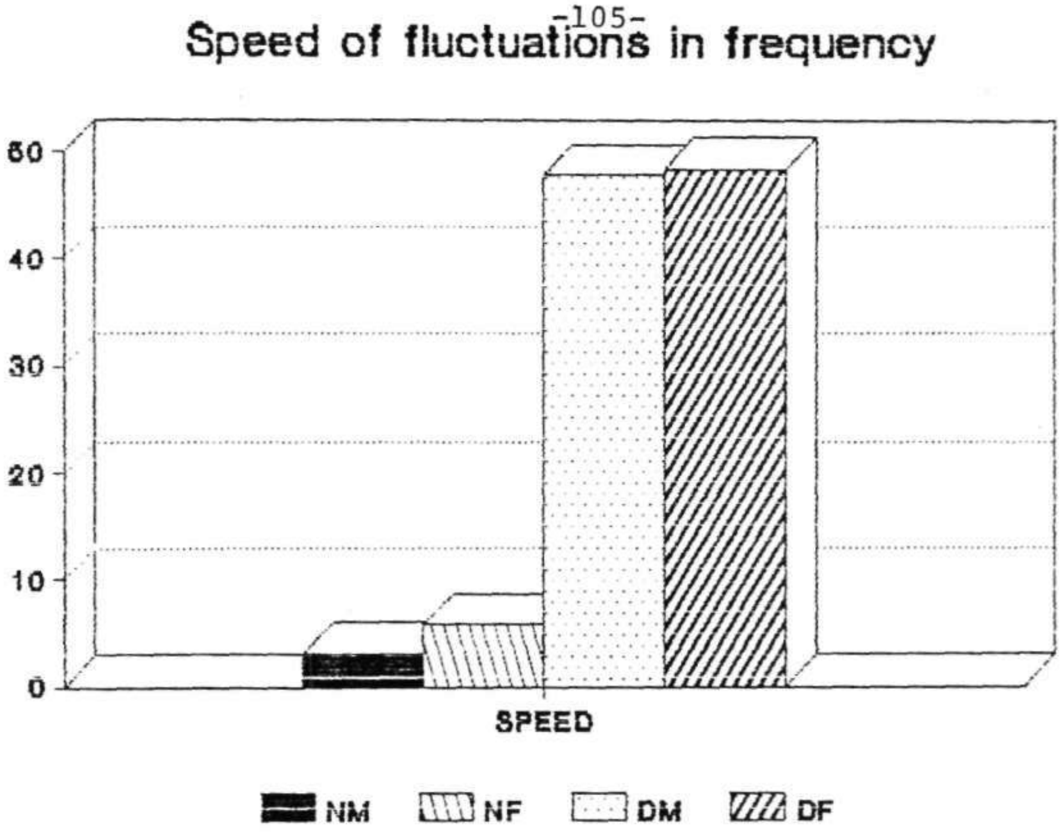
Groups	Mean	S.D.	Range
NM	3.0	4.12	0-10
DM	47.59	24.60	8-96
NF	5.6	0.92	8-12
DF	48.31	26.58	8-92

Table 7.1: The mean, S.D. and range of speed of fluctuations in fundamental frequency in normal and dysphonic males and females.

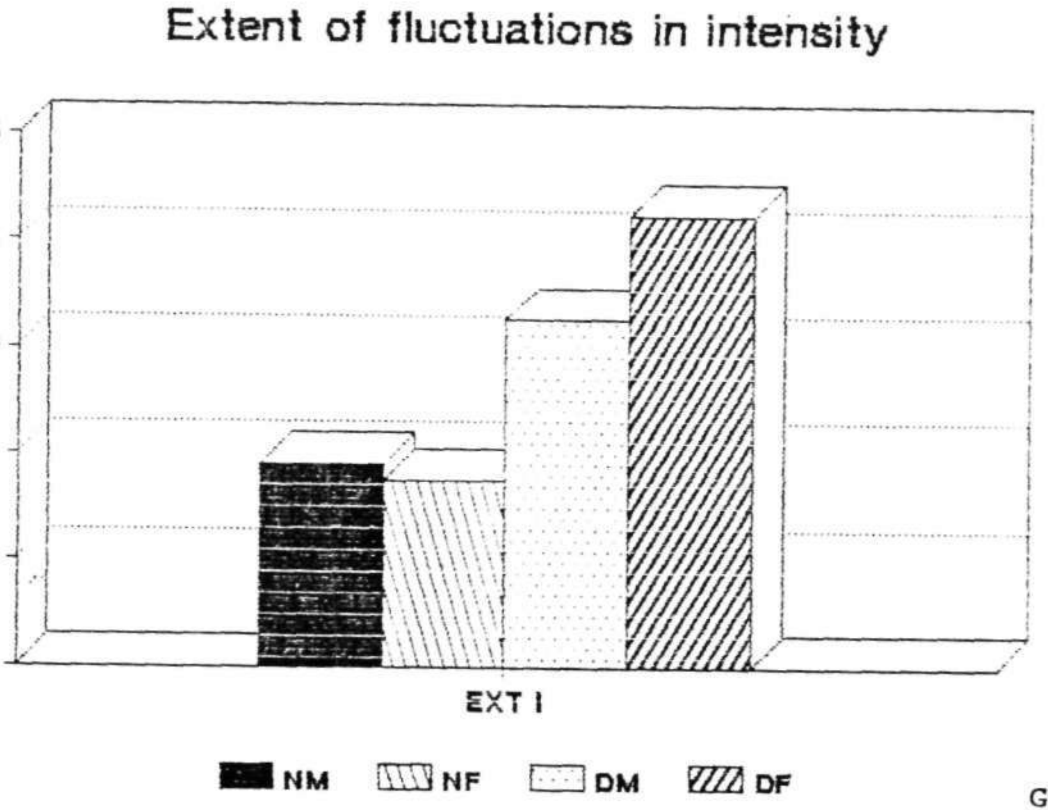
Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	

(+ significant at 0.05 level)

Table 7.2: Comparison of normal vs dysphonics, both males and females in terms of speed of fluctuation in fundamental frequency.



GRAPH 5



GRAPH 6

NM - Normal Male; NF - Normal Female; DM - Dysphonic Male;
DF - Dysphonic Female

Statistical analysis revealed a significant difference in this parameter between normal males and dysphonic males.

Hence the hypothesis stating that there is no significant difference between normals and dysphonics both males and females for speed of fluctuations in frequency is partially rejected and the hypothesis stating that there is no significant difference between

a. normal males and normal females

b. dysphonic males and normal dysphonic females for speed of fluctuations in frequency is accepted.

The findings of the present study are in confirmity with the study by Nataraja (1986) and this parameter could sufficiently differentially diagnose normals and dysphonics.

6. Extent of fluctuations in intensity

The values of the parameter for the males and females of both the groups are presented in tables given below (Tables 8.1 and 8.2) and in the Graph (Graph 6).

Groups	Mean	S.D.	Range
NM	1.9	1.76	0-3.6
DM	3.27	1.49	0-6.8
NF	1.75	1.94	0-4
DF	4.24	2.91	0-33.0

Table 8.1: The mean, S.D. and range of the extent of fluctuations in intensity for both the groups.

Group	Significance
NM vs NF	-
NM vs DM	-
DM vs DF	-
DF vs NF	-

(+ > significant of 0.05 level)

Table 8.2: Comparison of normal vs dysphonics, both males and females in terms of extent of fluctuations in intensity.

Study of tables and graph showed that the dysphonic males and females had slightly higher scores than normal males and females. Both normals and dysphonics did not show much variation.

Hence the hypothesis stating that there is no significant difference between

a. normals and dysphonics both males and females

b. normal males and normal females

c. dysphonic males and dysphonic females for extent of fluctuations in intensity.

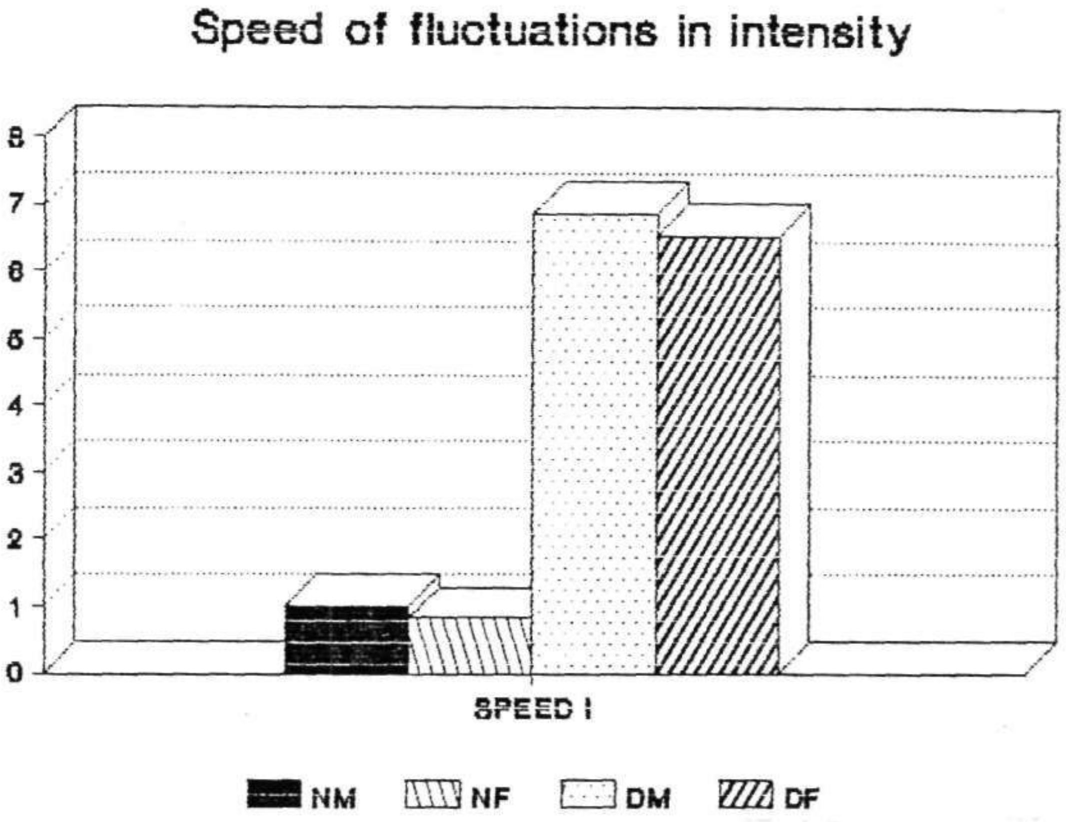
The statistical analysis revealed that there was no significant difference between normal and dysphonic groups. This is in confirmation with the studies reported by Kim et al., (1982) and Nataraja (1986). Hence the values of this parameter obtained in the study was similar to the previous studies. This parameter was not found to be very useful ni differentiating normals from dysphonics. However, according to Yoon et al. (1984) only a few advanced cases of carcinoma, had shown statistically significant difference in this parameter, when compared to normals. Hence this parameter may supplement information regarding the condition of the vocal cords along with other parameters.

7. Speed of fluctuation in intensity:

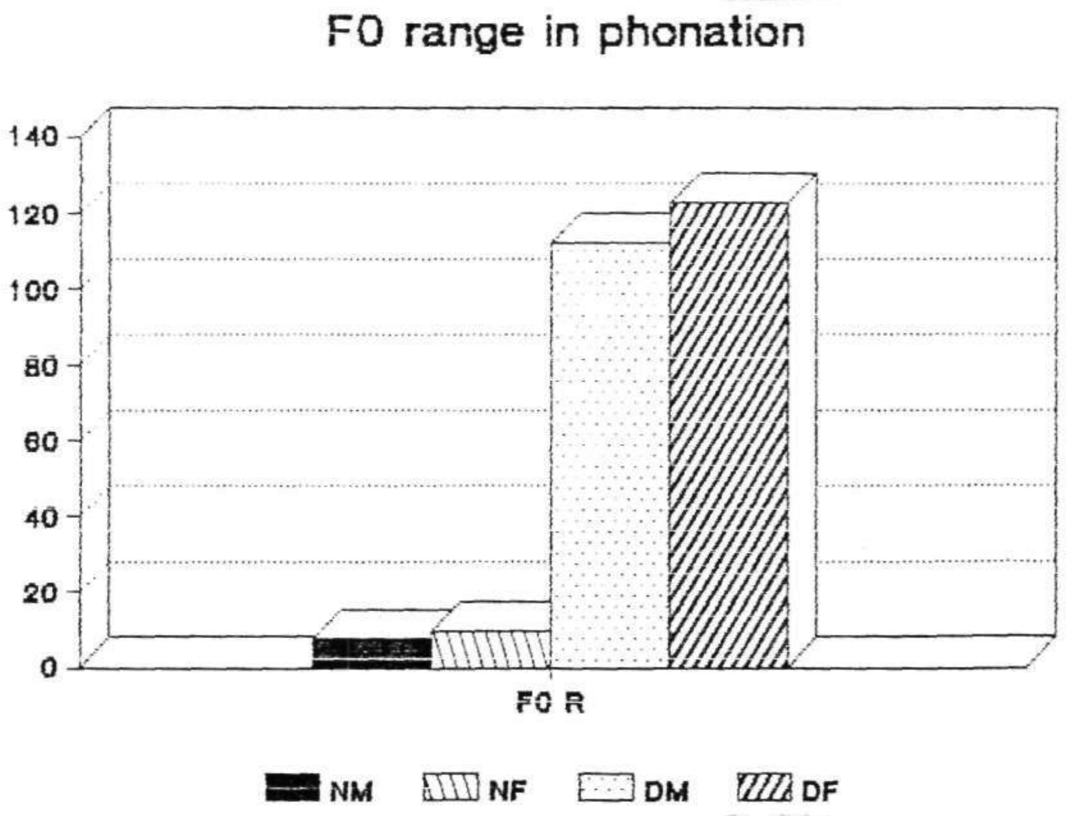
The values obtained for this parameter for both males and females of both groups are given in the tables (Tables 9.1 and 9.2) and shown in the Graph (Graph 7).

Groups	Mean	S.D.	Range
NM	1.0	1.0	0-2
DM	6.88	5.68	1-19
NF	0.83	0.93	0-2
DF	6.54	5.01	0-21

Table 9.1: The mean, S.D. and range in normals and dysphonics both males and females interms of speed of fluctuation in intensity.



GRAPH 7



GRAPH 8

NM - Normal Male; NF - Normal Female; DM - Dysphonic Male; DF - Dysphonic Female

Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	+

(+ significant of 0.05 level)

Table 9.2: Comparison of normal vs dysphonics, both males in terms of speed of fluctuations in intensity.

The study of the tables and graph indicated that the values obtained by dysphonic males and females were greater than that of normal males and females and no difference between males and females of both the groups were seen.

The statistical analysis showed that there was significant difference between normal and dysphonic males and females with this parameter.

Hence the hypothesis stating that there is no significant difference between normals and dysphonics both males and females for speed of fluctuations in intensity is rejected and the hypothesis stating that there is no significant difference between

a. normal males and normal females

b. dysphonic males and dysphonic females for speed of fluctuations in intensity.

These findings are consistent with the study by Nataraja (1986).

8. Fundamental frequency range in phonation:

The table showing the fundamental frequency range in phonation of male and females in both the groups are given below (Tables 10.1 and 10.2). The mean values of the parameter are given in the graph (Graph 8) also.

Groups	Mean	S.D.	Range
NM	7.80	4.79	1-16
DM	112.33	115.32	12-245
NF	9.43	6.89	3-29
DF	122.50	105.8	13-323

Table 10.1: The mean, S.D and range of fundamental frequency range in phonation for both normal and dysphonic groups.

Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	+

(+ significant of 0.05 level)

Table 10.2: Comparison of normal vs dysphonics in terms of fundamental frequency range in phonation.

Study of the tables and graph indicate that the dysphonic males and females had much higher values and greater variability when compared to their counterparts of normal group. It was noticed that there was not much significant difference between males and females of normals as well as dysphonic groups. These observations were further confirmed by statistical analysis, i.e. no significant difference between males and females were noticed both in dysphonics and normals.

A statistically significant difference was found between the normal and dysphonic groups with the males and females.

Hence the hypothesis stating that there is no significant difference between normals and dysphonics both males and females for frequency range in phonation is rejected and the hypothesis stating that there is no significant difference between

a. normal males and normal females

b. dysphonic males and dysphonic females for frequency range in phonation is accepted.

These result are in support with the study by Nataraja (1986) who report of similar findings. Hence the values obtained in this parameter were similar to those obtained in the previous study. This phonation was found to differentiate normal and dysphonic groups.

Intensity range in phonation

Groups	Mean	S.D.	Range
NM	3.80	2.8	1-8
DM	8.74	2.74	1-22
NF	3.66	2.06	1-7
DF	9.46	6.40	1-20

Table 11.1: Mean, S.D. and range of intensity range for phonation for both groups.

Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	+

(+ significant of 0.05 level)

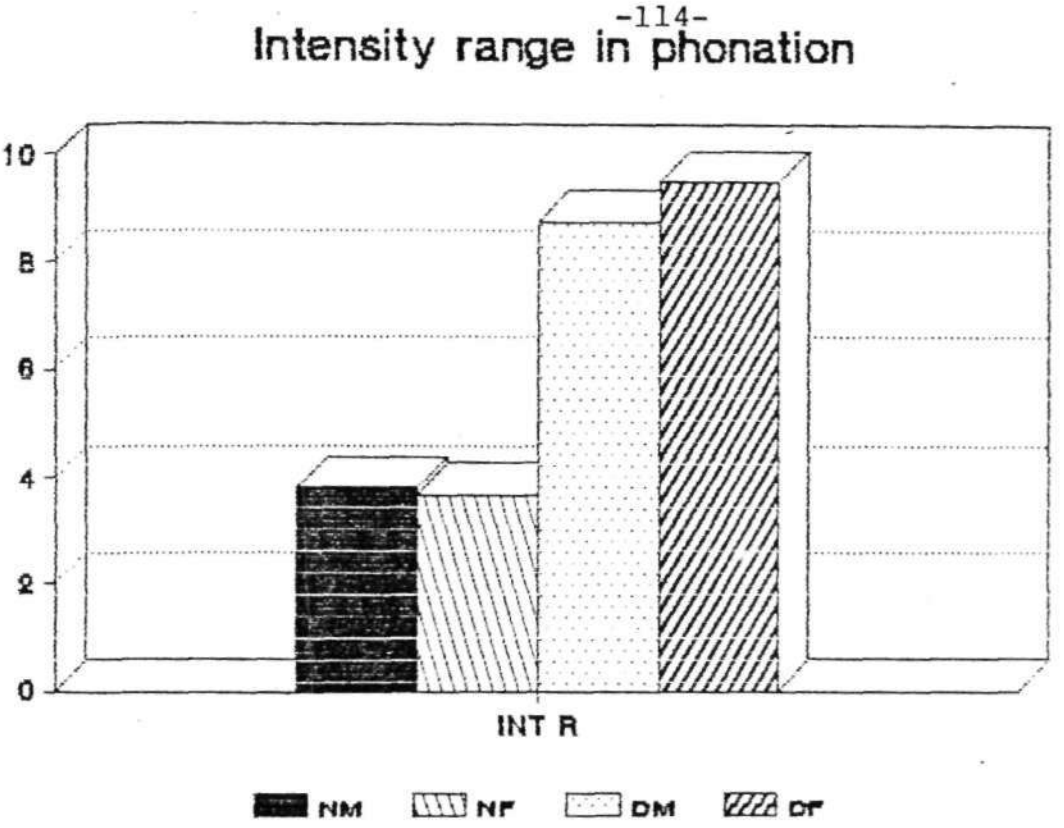
Table 11.2: Comparison of normals vs dysphonics, both males and females in terms of intensity range in phonation.

A statistically significant difference was observed between normal and dysphonic groups (for both males and females).

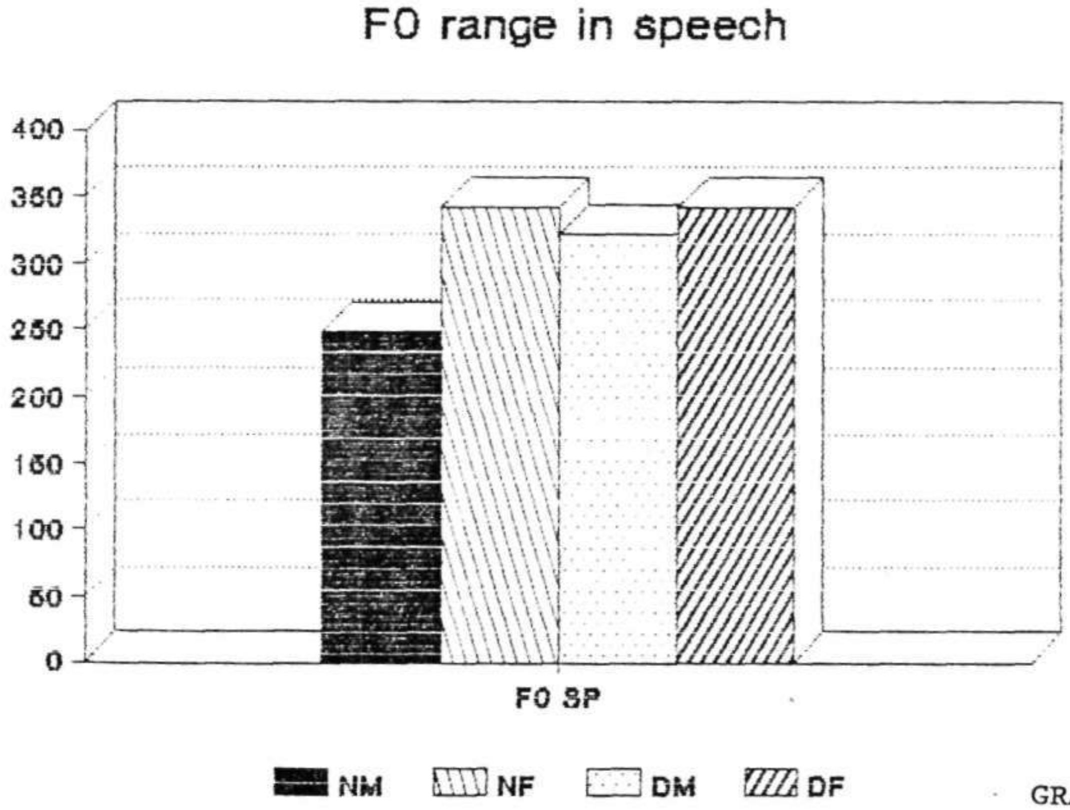
Hence the hypothesis stating that there is no significant difference between normals and dysphonics both males and females for intensity range in phonation and hypothesis stating that there is no significant difference for

a. normal males and normal females

b. dysphonic males and dysphonic females is



GRAPH 9



GRAPH 10

NM - Normal Male; NF - Normal Female; DM - Dysphonic Male; DF - Dysphonic Female accepted. Similar findings were also reported by Nataraja (1986).

10. Fundamental frequency range in speech

The fundamental frequency range in speech in normal and dysphonic males and females are shown in the tables (Tables 12.1 and 12.2).

Groups	Mean	S.D.	Range
NM	248.13	65.01	117-387
DM	321.56	87.38	121-470
NF	342.15	73.48	184-421
DF	343.12	61.5	225-496

Table 12.1: Mean, S.D. and range of fundamental frequency range in phonation for both groups.

Group	Significance
NM vs NF	-
NM vs DM	+
DM vs DF	-
DF vs NF	+

(+ significant at 0.05 level)

Table 12.2: Comparison of normals vs dysphonics, both males and females in terms of fundamental frequency in phonation.

Lowest mean was observed in normal males (248 Hz) and highest in dysphonic females (345 Hz) with dysphonic males and normal females (321 Hz and 342 Hz) falling in between the two.

In the dysphonic group no significant difference were found between males and females. Similarly no significant difference between males and females of normal group were seen. There was no significant difference between females of the two groups, and males of the two groups showed significant difference. All the groups had shown equal variability. Hence the values obtained in this parameter was similar to the previous study. This parameter was found to be useful in differentiating normals from dysphonics.

Similar finding were reported in the study by Nataraja (1986). Thus analysis of data on various parameters has shown that the following parameters were able to differentiate dysphonics both males and females from normals.

1. Mean air flow rate (MAFR)
2. Maximum phonation duration (MPD)
3. Fundamental frequency for phonation
4. Extent of fluctuations in frequency
5. Speed of fluctuations in frequency
6. Speed of fluctuations in intensity
7. Fundamental frequency range in phonation
8. Intensity range in phonation
9. Fundamental frequency range in speech

The following parameters showed significant difference between males and females of the normal group.

1. Maximum phonation duration
2. Fundamental frequency in phonation

None of the parameters showed significant difference between males and females of the dysphonic group.

Hence the values of the parameters taken up for the study were similar to the values obtained in the previous study (Nataraja, 1986). Hence the parameters chosen for the study could sufficiently differentiate normal and dysphonic groups.

As mentioned earlier before the data was fed into the data was preprocessed using a filter, to analyse to data in terms of their even distribution, skewness and kurtosis.

After the data was pre-processed using the filter the following fields were taken to be fed in to the nueral network, for processing and classfication.

1. Age of the subject
2. Sex of the subject
3. Fundamental frequency of phonation
4. Fundamental frequency range in phonation
5. Fundamental frequency range in phonation
6. Extent of fluctuations in frequency
7. Speed of fluctuations in frequency
8. Extent of fluctuations in intensity
9. Speed of fluctuations in intensity
10. Maximum Phonation Duration (MPD)
11. Mean Airflow Rate (MAFR)
12. Diagnosis of the Voice Disorder (DVD)

The data was trained using the Multi Layer Perceptron (MLP) Neural network model by using the data of 444 subjects. In order to obtain the best possible model to classify voice disorders the neural network was individually trained by adjusting the following parameters of the neural network (MLP).

1. The learning algorithms function of the neural network - The two learning algorithmic - Conjugate gradient and steepest descent - were taken up and the neural network was trained in each of these algorithms to find which of the learning algorithm gave the best results.

2. The number of hidden layers - The network was trained with one and two hidden layers and it was calculated as to which option could give the best results.

3. The number of nodes in each hidden layer - The number of nodes in each hidden layer was adjusted sequentially from 1 to 10 in the case of the 1st hidden layer and 1 to 8 in the case of the 2nd hidden layer.

The results indicated the following.

1. Using the learning algorithm 'Conjugate gradient' resulted in poor overall results when compared with the learning algorithm 'steepest descent'.

2. By increasing the number of hidden layers to two, the neural network could classify the data samples more effectively than by using 1 hidden layer.

3. The best performance was obtained by using 6 nodes in the 1st hidden layer and 3 nodes in the 2nd hidden layer in the case of 'steepest descent' learning algorithm function

and using 8 nodes and 3 hidden layers in the case of conjugate gradient.

By using 6 nodes & 3 nodes in the hidden layer in the learning algorithm of steepest descent the network model could correctly predict the voice disorder by 88.10%, where as by using 8 nodes & 3 nodes in the 1st and 2nd hidden layers in the learning algorithm 'conjugate gradient' the network could correctly predict the voice disorder only by 70.33%. Hence using the algorithm of steepest descent with 6 nodes in the 1st hidden layer and 3 nodes in the second was the best model which could classify normal subjects and subjects with voice disorders and which could also classify the voice disorders.

During the training phase a total of 333 datas were trained. The results of the data are given in the form of cross tabulation matrix as shown below.

	e	c	b	a
e	51	2	2	0
c	2	129	1	0
b	9	1	121	0
a	0	0	0	15

It can be seen from the matrix that out of a total of 333 data, 316 of them were identified correctly. In terms of percentage the network could identify the various subgroups 94.89% of the times. In terms of the individual pathologies, out of the total of 131 subjects who had mass pathologies the network correctly identified 121 of them. It identified 9 of the mass pathologies as congestion and 1 of the mass pathologies as glottic chink.

Out of the total of 132 subjects who had glottic chink as their pathology, the network correctly identified 129 of them. 2 of the subjects having glottic chink was identified as congestion and one of the subject with glottic chink was identified as a case with mass on the vocal folds.

Out of the total of 59 subjects who had congestion as the pathology in vocal folds, the network correctly identified 51 of them. 2 of the subjects with congestion were identified as subjects with mass and two of the subjects with congestion were identified as having glottic chink.

Out of the 15 normal subjects in the training data, the network correctly identified all of them as normals.

The identification scores of the network in terms of percentages for the normal and dysphonic subjects are shown below in the table [table 13].

Subjects	Total No. of Samples	Samples Correctly identified	Percentage
Normal	15	15	100%
Mass	131	121	92.37%
Glottic Chink	132	129	97.73%
Congestion	59	51	86.44%

Table 13: percentage of the number of subjects identified correctly by the Neural network in the training phase.

In the validation data total of 41 samples were taken. The results of the data are given in the form of cross tabulation matrix as shown below.

		Predicted			
		e	c	b	a
True	e	7	0	0	0
	c	0	15	3	0
	b	0	3	11	2
	a	0	0	0	9

a = Normal
b = Mass
c = Glottic Chink
e = Congestion

Total number = 50
Number of correct responses = 42
Percentage of correct responses = 84%

The study of table shows that from a total data of 50 subjects in the validation data, the network could identify correctly 42 of the subjects.

In terms of the individual pathologies, out of the 16 subjects with mass on the vocal folds, the network was able to correctly identify 11 of them. 3 subjects with mass were identified as glottic chink and 2 subjects with mass were identified as normals.

Out of 18 subjects with the pathology of Glottic chink, the network identified 15 of the subjects correctly. 3 subjects with Glottic chink were identified as mass by the network.

Out of 7 subjects with congestion in the vocal folds, the network correctly identified all the 7 subjects.

In the validation data there were atotal of 9 normals, again as in the training data, the network was able to identify all the 9 subjects correctly.

The table below (Table 14) shows the distribution of the subjects and the percentage of correct identification by the neural network.

Subjects	Total Samples	Samples correctly identified	Percentage
Normals	9	9	100%
Mass	16	11	68.75%
Glottic chink	18	15	83.33%
Congestion	7	7	100%

Table 14: Percentage of the number of subjects identified correctly by the neural network in the validation phase.

The test data consisted of 45 samples. The results of the test data are shown in the cross tabulation matrix.

		Predicted			
		e	c	b	a
True	e	6	0	0	0
	c	2	16	0	0
	b	1	2	12	0
	a	0	0	0	6

a = Normal
b = Mass
c = Glottic Chink
e = Congestion

Total number = 45
Number of correct responses = 40
Percentage of correct responses = 88.80%

From the table it can be seen that 88.10% out of a total of 45 subjects 40 were identified correctly. In terms of percentage the network was able to correctly identify the pathologies 88.10% of the times.

In terms of data of individual pathologies out of the total of 15 subjects with a mass pathology, the network was able to identify 12 subjects correctly 2 subjects with mass on vocal folds were identified as having glottic chink and 1 subject with mass pathology was identified as having congestion of vocal folds.

Out of 18 subjects with glottic chink the network identified 16 subjects correctly 2 of the subjects with glottic chink were identified as congestion. Out of 6 subjects with congestion, the network identified all the 6 subjects correctly. In the normal group, 6 normal subjects were taken for the training data, and all the 6 subjects were correctly identified.

The distribution of subjects and the percentage of correct identification by the neural network is shown in the table below.(Table 15)

Subjects	Total Samples	Samples correctly identified	Percentage
Normal	3	3	100%
Mass	15	12	80%
Glottic chink	10	16	88.8%
Congestion	6	6	100%

Table 15: Percentage of the number of subjects identified correctly by the neural network in the test phase.

Finally the run data consisted of 12 subjects, 2 from each group. The results of the run data are shown in the cross tabulation matrix.

	e	c	b	a
e	2	0	0	0
c	0	1	1	0
b	0	0	1	1
a	0	0	0	0

a = Normal
b = Mass
c = Glottic chink
e = Congestion

Total number = 8
Number of correct responses = 6
Percentage of correct responses = 75 %.

From the matrix it can be seen that out of a total of 8 subjects the network could correctly identify all the normals and all the subjects with congestion. Out of 2 subjects with mass in the vocal folds, the network could correctly identify 1 of them. The other was identified as normal. Out of 2 subjects with glottic chink the network correctly identified 1 of them and the other was identified as mass on the vocal folds. The distribution of subjects and the percentage of correct of correct identificationby the network is shown in the table below. (Table 16)

Subjects	Total samples	Samples correctly identified	Percentage
Normals	2	2	100%
Mass	2	1	50%
Glottic chink	2	1	50%
Congestion	2	2	100%

Table 16: Table showing percentage of subjects correctly identified During the run phase.

From the data discussed above, it can be seen that in the stages of training, validation,test and the run data, the neural network could correctly identify all the normal subjects and the subjects with congestion in vocal folds, where as the correct identification of subjects with mass in the vocal folds or subjects with glottic chink were was comparatively poor. In subjects with mass in their vocal folds, during the training phase the network could identify a mass pathology 92.37% correctly, where as in the validation phase the correct identification was reduced to 68.75% and in the testing phase the performance improved, and the percentage of correct identification was 80%.

For subjects with glottic chink during the training phase the network could identify the subjects correctly 97.73% of tines, where as in the validation phase the percentage of correct identification reduced to 83.3% and in the test phase the percentage of correct identification was 88.8%. This variation was basically because of the values of parameters obtained by different subjects involved in different stages.

The network could identify all the normal subjects correctly during the training, test and run phases (100%). This result is in agreement with the previous studies which report of significant difference between normal and dysphonic voices using acoustic and aerodynamic parameters. (Nataraja,1986; Yoon et., al,1984; Kim et.,al,1982; Jayaram,1975; Shigemori,1977)

In the training data 9 of the subjects with mass pathologies were identified as congestion, 1 of them was identified as having glottic chink.

2 of the subjects who had glottic chink were identified as congestion and 1 was identified as having mass on the vocal folds. 2 of the subjects with congestion was identified as subjects with mass on the vocal folds and two of the subject were identified glottic chink. In the validation data 3 subjects with mass on vocal folds were identified as glottic chink and 2 subject with mass on the vocal folds were identified as normals. 3 subject with glottic chink were identified as mass on the vocal folds.

In the test data it was seen that 2 subject with mass on the vocal folds were identified as glottic chink and 1 subject with mass on the vocal folds was identified as having congestion. 2 of the subjects with glottic chink was identified as congestion.

In the run data 1 subject having mass on the vocal folds was identified as normal and one subject having glottic chink was identified as having mass.

Two of the subjects In the training phase and 1 subject in the run phase with mass on the vocal folds were identified as normal. Two of these subjects had an 'early vocal nodule' and other had mild thickening. The variation in the acoustic and aerodynamic parameters from normal subjects may be very minimal and hence estimation of the pathology from the acoustic and aerodynamic data could be difficult, hence the neural network had not been able to distinguish between normal subjects from dysphonics. It is also seen that subjects with mass on the vocal folds were identified as glottic chink and vice versa by the nueral network. This could be because of the fact that the glottic chink was characterized by escape of air from the glottis during phonation and in subjects with mass on the vocal folds also due to the presence of growth there was mostly escape of air through the glottis. Hence in subjects with mass and glottic chink the acoustic and aerodynamic parameters were similar. Hence the classification of a mass and a congestion pathology could be difficult from acoustic or aerodynamic parameters. But the subjects with congestion had no air escape from the glottis during phonation when compared to mass on the vocal folds and pathologies with glottic chink.

Hence congestion can be classified distinctively from a mass on the vocal folds or glottic chink.

However the neural network proved to be capable of classifying the normals and dysphonics based on the aerodynamic and acoustic paramaters that were used. Further it has also been seen that it was capable of distinguishing different types of laryngeal pathologies based on the parameters used with high degree of reliability and validity. It has also been showed by the present experiment that it is possible to use the neural network to classify normals and dysphonics and different types of dysphonias. The validity and reliability of the classification of different types of disorders would depend on the parameters used and also the number of subjects that are used for training the network. Better results can be obtained by increasing the number and varieties of cases. Further overlapping or borderline cases can also be classified by adding more sensitive parameters like EGG parameters or perturbations parameters.

Thus the objective of the study to classify the normals and dysphonics and differentiation of dysphonics based on 10 aerodynamic and acostic parameters using artificial neural networks has been achieved. The study has opened up new possibilities of computer applications in the field of Speech and Hearing.

SUMMARY & CONCLUSION

The present study was desinged to objectively classify voice disorders using an artificial neural network. Voice disorders in terms of the acoustic and aerodynamic parameters using an artificial nueral network.

Based on the review of literature the following acoustic and aerodynamic parameters were taken up for the study.

Aerodynamic Parameters

1. Mean air flow rate (MAFR)
2. Phonation Duration (PD).

Acoustic Parameters

1. Fundamental frequency in phonation.
2. Extent of fluctuations in frequency.
3. Speed of fluctuations in frequency.
4. Extent of fluctuations in intensity.
5. Speed of fluctuations in intensity.
6. Fundamental frequency range in phonation.
7. Intensity range in phonation.
8. Fundamental frequency range in speech.

The total of 444 subjects were Considered for the study out of which 414 were dysphonic subjects and 30 of them were normal subjects.

The voice pathologies included in the present study included the following

a. Man on the vocal folds (including vocal nodule, vocal polyp and thickening)
b. Glottic chink (including subjects with glottic chink and paralysis of vocal folds).
c. Congestion.

The normals taken up for the study had an age range of 20-25 years and the dysphonics taken up for the study had an age range of 25-75 years.

The normal group (N=30) consisted of 15 males and 15 females. The dysphonic group (N-414) consisted of 274 males and 140 females.

The input data consisting of acoustic and aerodynamic parameters and the output data consisting of the diagnosis of voice disorders were fed to the neural network. The neural network used for this study was Multi Layer Perception (MLP). After training the data, the Nueral network model was tested on the test data and it was found that the neural network could classify normal and abnormal voice and categize the voice disorders correctly. (88.10%)

Within the subject categories it was seen that normals subjects and dysphonics with congestion were identified better than dysphonics with mass On the vocal folds and glottic chink.

CONCLUSIONS

1. The neural network can be trained using ten aerodynamic and acoustic parameters to classify normals and dysphonics.
2. Differentiaton of different types of dysphonia was also possible using an artificial neural network.
3. Ten aerodynamic and acoustic parameters were capable of classifying 4 types of dysphonias (88.1%) and differentiate Dysphonics from normals.
4. The reliability and validity of the classification could be improved by altering the hidden layers and nodes of the neural network.
5. Efficiency of the system could be improved more and better data.

Thus the objective of classifying the different types of dysphonia was achieved using the artificial neural network.

BIBLIOGRAPHY

ARNOLD, G.E. (1955) : "Vocal rehabilitation of paralytic dysphonia". Acoustic analysis of vocal fold function. Archives of otolarynyology., 52: 593-601.

Asthana, P. (1977) : "Relationship between vocal intensity, pitch and Nasality in cleft palate speakers". Unpublished Master's Dissertation; University of Mysore.

Baer, T. (1980) : "Vocal jitter - A neuromuscular expanation". Transcript of the eight symposium of the care of the professional voice, Voice foundation - New York (19-22).

Berry, M.F., and Eisenson, J. (1962) : "Speech Disorders: principles, practice of therapy", Peter Owen Ltd., London.

Bohme, G., and Hecker, G. (1970) : "GERON TOLOGISCHE UNTERSUCHUNGEN UBER STIMMUMFANG UNDSPERCHSTIMMLAGE". Folia Phoniatrica., 22:176-184.

Boone,D.E. (1983) : "The voice and voice therapy", (2nd edition), Prentic - Hall Inc., Engle Wood Cliffs, N.J.

Bouhuys,A., Proctor,D.F., and Mead, J. (1966) : "Kinetic aspects of singing". Jounal of applied Physiology; 21:483-496.

Bowler,N.W. (1964) : "A fundamental frequency analysis of harsh vocal quality". Speech Monograph; 31: 128-134.

Hippel, K. and McOwinski, D. (1978) : "Examination of patients with normal and abnormal voice by pneumatachography". H.No: 26(12) : 421-423.

Hirano, M. (1981) : "Clinical examination of voice". Disorders of human communication, 5, springer, Wien.

Hirano, M., (1975): "Phono surgery-Basic and clinical investigations". Otologia (Fukuoka); 21: 239-440.

Hirano, M. Kioke, Y.J. and Vonleden, H. (1968) : "Maximum phonation time and air usage during phonation" : a clinical study. Folia phoniatrica ;20: 185-201.

Hirano, M. Miyahara, T., Hirose, H., Kiritani. S., and Fujiyama, J . (1970) : "Air flow rate in singing. Annual bulletin, Research Institute of Logopedics and Phoniatrics", University of Tokyo., 4: 55-56.

Horii. Y. (1979) : "Fundamental frequncy perturbation observed in sustained phonation". Journal of Speech and Hearing Research : 22: 5-19.

Hudson. A.R., and Halbrook, A., (1981) : "A study of the reading fundamental vocal frequency of young black adults". Journal of Speech and Hearing Research., 24 (2) : 197-201.

Imaizumi, S., Hiki, S., Hirano, M., and Matushita, H. (1980) : "Analysis of Pathological Voices with sound spectoograph" Journal of Acoustical Society of Japan, 36 : 9-16.

BIBLIOGRAPHY

ARNOLD, G.E. (1955) : "Vocal rehabilitation of paralytic dysphonia". Acoustic analysis of vocal fold function. Archives of otolarynyology., 52: 593-601.

Asthana, P. (1977) : "Relationship between vocal intensity, pitch and Nasality in cleft palate speakers". Unpublished Master's Dissertation; University of Mysore.

Baer, T. (1980) : "Vocal jitter - A neuromuscular expanation". Transcript of the eight symposium of the care of the professional voice, Voice foundation - New York (19-22).

Berry, M.F., and Eisenson, J. (1962) : "Speech Disorders: principles, practice of therapy", Peter Owen Ltd., London.

Bohme, G., and Hecker, G. (1970) : "GERON TOLOGISCHE UNTERSUCHUNGEN UBER STIMMUMFANG UNDSPERCHSTIMMLAGE". Folia Phoniatrica., 22:176-184.

Boone,D.E. (1983) : "The voice and voice therapy", (2nd edition), Prentic - Hall Inc., Engle Wood Cliffs, N.J.

Bouhuys,A., Proctor,D.F., and Mead, J. (1966) : "Kinetic aspects of singing". Jounal of applied Physiology; 21:483-496.

Bowler,N.W. (1964) : "A fundamental frequency analysis of harsh vocal quality". Speech Monograph; 31: 128-134.

Brackett, I.P. (1971) : "Parmeters of voice quality" in Travis, L.E. (1971), (ed.) Hand Book of Speech Pathology and Audiology, Appleton Century - Crofts, N.Y.

Brodnitz, F.S., (1959) : Vocal Rehabilitation, Rochester, Minn., Whiting Press.

Carhart. R., (1938) : "Infra glottal resonance and a cushion pipe". Speech Monograph; 5 : 65-90.

Carhart, R., (1941) : "The Spectra of Model larynx tones". Speech Monograph; 8 : 76 - 84.

Chandrashekar, K.R., (1987) :"Electroglottography in dysphonics". Unpublished Master dissertation, A.I. I.S.H., University of Mysore.

Coleman, R. (1969) : "Effect of Median Frequency Levels upon roughness of filtered stimulti". Journal of Speech and Hearing Research; 12 : 330-336.

Coleman, R. and Wendahl (RW), (1967) : "Vocal Roughness and stimulus duration". Speech Monograph; 34 : 85-92.

Coombs, J. (1976) : "The Maximum Duration of Phonation of]a[in normal and hoarse - voiced children". Unpublished Master's thesis, Portland State University.

Cooper, M. (1974) : "Spectographic Analysis of Fundamental Frequency and Hoarseness before and after vocal rehabilitation". Journal of Speech and hearing disorders, 39 : 286-295.

Crystal and Jackson (1970) : "Methodology and results on Laryngeal disorder detection through spectral analysis", Final report, Contract pH-86-68-192, Singatron. Inc. Lexington, Massachusetts.

Darby, J.K. (Ed) (1981) : Speech Evaluation in Psychiatry, Grune and Stratton, Inc., N.Y.

Darely. F.L. (1964) : "Diagnosis and appraisal of communiction disorders". New Jersey, Prentice - Hall Engle Wood Cliffs.

Davis, S.B. (1976) : "Computer Evaluation of laryngeal Pathology based on inverse filtering of speech". SCRL Monograph, 13, California.

Eberhart, R.C., and Dobbins, R.W. (1994) : "Nueral Networks - A comprehensive foundation". IEEE; New York.

Emerick. L., and Hatten, J. (1974) : "Diagnosis and evaluation in speech pathology". New Jersey, Prentice - Hall. Englewood Cliffs.

Fair Banks, G. (1960) : "Voice and articulation drill book", New York. Harper and Row publishers.

Fant, G. (1960) : "Acoustic theory of speech production" : S.gravehage, Mounton and Co.

Farely, R.G. (1994) : "Controll of voice to Fo by an artificial neural network". Journal of Acoustic Society and America; 105 : 1411-1422.

Fisher, H.B. (1996) : "Improving voice and Articulation" Hough on Miffin, Co., Boston, New York.

Fitch, J.L. and Holbrook, A. (1970) : "Modal vocal fundamental frequency of young adults". Archives of otolaryngology; 92: 744-746.

Gopal. H.S. (1980) : "Relationship for locating optimum frequency in the age range of 7 to 25 years". Unpublished Master's Dissertation. University of Mysore.

Gould. W.J., (1975) : "Quantitative assetsment of voice function in microlaryngology". Folia Phoniatrica., 27(3) : 193-200.

Gould. W.J. and Okamura, H. (1974) : "Respiratory training of the singer". Folia phoniatrica; 26: 275-286.

Gary, G.W. and Wise, C.M. (1959) : " The bases of speech' Harper and Row publishers, New York.

Haykin, S. (1994) : "Neural Network : A comprehensive Foundation" New York, Macmillan.

Higgins, B.M., and Saxman, H. J. (1989): "A comparsion ci intra subject variation across sessions of three vowel frequency perturbation indices" Journal of Acoustic Society of America : 86 (3) : 911-916.

Heiberger, V.L. and Horii. Y. (1982) : "Jitter and Shimmer in sustained Phonation" in Lass, N.J., (ed) speech and Language, Vol.7 Academic press, New York.

Hippel, K. and McOwinski, D. (1978) : "Examination of patients with normal and abnormal voice by pneumatachography". H.No: 26(12) : 421-423.

Hirano, M. (1981) : "Clinical examination of voice". Disorders of human communication, 5, springer, Wien.

Hirano, M., (1975): "Phono surgery-Basic and clinical investigations". Otologia (Fukuoka); 21: 239-440.

Hirano, M. Kioke, Y.J. and Vonleden, H. (1968) : "Maximum phonation time and air usage during phonation" : a clinical study. Folia phoniatrica ;20: 185-201.

Hirano, M. Miyahara, T., Hirose, H., Kiritani. S., and Fujiyama, J . (1970) : "Air flow rats in singing. Annual bulletin, Research Institute of Logopedics and Phoniatrics", University of Tokyo., 4: 55-56.

Horii. Y. (1979) : "Fundamental frequncy perturbation observed in sustained phonation". Journal of Speech and Hearing Research : 22: 5-19.

Hudson. A.R., and Halbrook, A., (1981) : "A study of the reading fundamental vocal frequency of young black adults". Journal of Speech and Hearing Research., 24 (2) : 197-201.

Imaizumi, S., Hiki, S., Hirano, M., and Matushita, H. (1980) : "Analysis of Pathological Voices with sound spectoograph" Journal of Acoustical Society of Japan, 36 : 9-16.

Isshiki, N., and Vonleden, H. (1964) : "Hoarseness : Aerodynamic Studies". Archives of Otolaryngolgy,80 : 206 - 213.

Isshiki, N., Okamura, H., and Horimoto, M. (1967) : "Maximum Phonation time and airflow rate during phonation : Simple clinical tests for vocal function". Annals of Otology, Rhinology and Larynyology, 76 : 998-1007.

Iwata, S., and Vonleden, H. (1970) : "Phonation quotient in patients with laryngeal disease" Folia Phoniatrica: 22, 117-128.

Iwata, S., Vonleden, H., and Williams, D. (1972) : "Air flow measurements during phonation". Journal of Commuication disorders ; 5: 67-69.

Iwata, S., and Vonlden, H. (1970) " "Voice prints in laryngeal disease". Archives of Otolaryngology; 91 : 346-351.

Iwata, S. (1972) : "Periodicities of pitch perturbations in normal and pathological larynges". Laryngoscope : 82; 87-96.

Iwata, S., Esaki, T., Iwami, K., and Takasu, T. (1976) : "Laryngeal function in patients with laryngeal polyps after laryngomicrosurgery". Pract. otol. (Kyoto) : 69 : 499-506.

Jain, S.K., and Ramaiah, T. J. (1969) : "Normal Standards for pulmanory function tests for healthy, Indian men 15-40 years old". comparison of different regression equation (prediction formulae). Indian Journal of Medical Research: 57 : 1453-1466.

Jayaram, K. (1975) : "An attempt at different diagnosis of dysphonia". Master's dissertation, University of Mysore.

Kitajima (K) and Gould (WJ) (1976) : "Vocal Shimmer in sustained phonation of normal and pathologic voice". American Journal of Otology; 9 : 25-32.

Koike, Y. (1969) : "Vowel amplitude modulations in patients with laryneal diseases". Journal of acoustic society of America, 45 : 839-844.

Koike, Y. (1973) : "Application of some acoustic measures for the evaluation of laryngeal dysfunction". Studia phonologica; 7 : 17-23.

Komiyama, S. and Buma, S. and Watanbe, H. (1973) : "Measurement of the phonation time taking account of the intensity of the voice". Otorhinolarygology. Fukuoka; 19(3) : 385-389.

Kosko, B. (1992) : "Nueral Netowrks and Fuzzy System" - A dynamic system approach to machine intelligence. Prentic Hall, Engle Wood Cliffs, N.J.

Krishnan, B.T. and Vareed, c. (1932) : "The vital Capacity of 103 medical students in South India". Ind. J. Med. Res 19 : 1165 - 1183.

Krishnamurthy, B.N. (1986) : "The Measurement of mean air flow rate in normals". Unpublished Master's dissertation. University of Mysore.

Kunze, L.E. (1964) : "Evaluation of Methods of estimating subglottal air pressure" : Journal of Speech and Hearing Research ; 151-164.

Kim, K.M. and Kakita, Y. and Hirano, M. (1982) : "Sound spectrographic analysis of the voice of patients with recurrent laryngeal nerve paralysis". Folia Phoniatrica, 34 : 124 - 133.

Large, J. Iwata, S. and Leden, H. Von. (1972) : "The male operatic head register versus falsetto". Folia phoniatrica; 24 : 19.

Lass, N. and Michel, J. "The effet of frequency, Intensity and vowel type on the maximum duration of phonation". Univ. Kan. Unpublished manuscript.

Lass, N. J. Brong, G. W., Ciccolella, S. A., Walters, S. C, and Maxwell, R.I. (1980) : "An investigation of speaker height and weigt discriminations by means of paired comparison judgements. Journal of phonetics; 8 ( )

Launer, P.G. (1971): "Maximum Phonation time in Children". Thesis state University of N.Y. Buffalo (Unpublished).

Lewis, K., Casteel, R., and Mc. Mohan, J. (1982) : "Duration of Sustained |a| related to the number of trials". Folia phoniatrica; 34(1) : 41-48.

Leinonen, L., Hihunen, J. Laakso, M.L., Rihkanen, H. and Pop/ius, H. (1997) : "Categorization of voice disorders with six perceptual dimensions". Folia phoniatrica; 49 : 9-20.

Leinonen, L., Kangas, J. Torkkola, K. and Juvas, A. (1992) : "Dysphonia detected by pattern recognition of spectral composition". Journal of Speech and Hearing Research; 35 : 287-295.

Leinonen, L., Mujunen, R., Kangas, J. and Torkkola, K. (1993) : "Acoustic pattern recognition of frictive - vowel coartiuclation by the self-organizing map". Folia Phoniatrica 45 : 173-181.

Lieberman (P) (1961) : "Perturbations in vocal pitch". Journal of Acoustic socieity of America; 33 : 597-603.

Lieberman (P) (1963) : "Some measures of the fundamental periodicity of normal and pathological larynges". Journal of Acoustic Society of America; 35 : 344-353.

Luchsinger, R and Arnold, G.E. (1965) : "Voice - Speech Language - Clinical Communicology" : It's Physiology and Pathology. Constable and Co. Ltd.

Ludlow, C, Coulter, D., and Cardano, C. (1979) : "Application of pitch perturabtion measures to the assessment of hoarsensess in Parkinson's disease". Jorunal of Acoustic Society of America; 66 : 565-566.

Mandi, A., Lengyal, E., and Csukas, M. (1976) : "Evaluation of respiratory function tests in aged persons". Excerp. Med. Che. Disorders; 30 : (1 -).

Markel, J.D. (1973) : "Application of a digital inverse filter for automatic formant and Fo anlaysis". IEEE transactions on Audio and Electroacoustics; AU-21 : 154-160.

Mc. Glone, R.E. (1967) : "Air flow during vocal fry phonation". Journal of Speech and Hearing Research ; 10 : 299-304.

Mc Glone R.E. and Hollien, H. (1963) : "Vocal Pitch Characteristics of aged women". Jorunal of Speech and Hearing Research; 6 : 164-170.

Meenakshi, S. (1984) : "Spirometric Data in South Indian Subjects, 60 to 80 years of age". Ind. J. of Physio, and Pharmacol; 28 : 58-62.

Menon, D. (1996). "The analysis of voice before and after treatment". Unpublished Master's dissertation, University of Mysore.

Metz. D.E., Schiavetti, N. and Knight, S.D. (1992) : "The use of artificial neural networks to estimate speech intelligibility from acoustic variables". A preliminary analysis. Journal of Communication Disorders. 25 : 43-53.

Michel, and Wendahl, R. (1971) : "Correlates of voice production" in TRAVIS, L.E. (ed) Hand book of Speech Pathology and Audiology Precentice - Hall, Inc., Englewood Cliffs, N.J. 465-480.

Moore (P) and Von Leden (H). (1958) : "Dynamic variations of the vibratory pattern in the normal larynx". Folia Phoniatrica; 10 : 205 - 238.

Markel, N., Meisels, M. and Houck, J. (1964) : "Judging Personality for voice quality", Jr. abnorm - Soc. Pschol., 69 : 458-463.

Mujunen, R., Leinonen, L., Kangas, J. and Tookkola, K. (1993) : "Acoustic pattern recongition of)sj misatriculation by elf-organizing map". Folia phoniatrica 45. 135-144.

Murry, T. (1978): "Speaking Fundamental Frequency Characteristics Associated with voice pathologies", Journal of Speech and Hearing Dissoders; 43(3) : 374-379.

Murry, T. and Doherty, E.T. (1980) : "Selected Acoustic Characteristics of pathologic and normal speakers". Journal of Speech and Hearing Research; 23(2) : 361-369.

Nag. S.K., Chatterjee, S. and Dey, S.K. (1982) : "Influence of long term Cigaratte smoking on ventilatory measurements of Calcutta People" I.J.P.P.A.Z., 26 : (34-).

Nadaleczny, M and Luschinger, R. (1934) : "Vitalkapazitat and Atem Typers bie sangern". Schw. Med. Wschr; 19 : (453).

Nataraja, N.P. (1972) : "Objective method of locating optimum pitch". Unpublished Master's Dissertation. University of Mysore.

Nataraja, N.P.(1984 b) : "Mean air flow rate and optimum frequency". Jorunal of All India Institute of Speech and Hearing;15 : 133-136.

Nataraja, N.P and Jagadish, A. (1984) : "Vowel Duration and fundamental freqeuncy". Journal of All India Institute of Speech and Hearing ; 15 : ( ).

Nataraja, N. P and Jayaram, M. (1982) : "A new apprach to the classification of voice disorders". Journal of All India Institute of Speech and Hearing; 8 : 21-28.

Nataraja, N.P. and Rashmi, M. (1984) : "Vital Capacity and its relation to height and weight". Journal of All India Institute of Speech and Hearing ; 15 : 53-56.

Nataraja, N.P. (1986) : "Differential Diagnosis of Dysphonias", Doctoral Thesis, A.I.I.S.H., University of Mysore.

Nigrin, A. (1993) : "Neural netowrks for pattern recognition", Cambridge, MA : The MIT Press.

Patacek, P.H. and Sader, E.K. (1963) : "Maximum duration of phonation". Jorunal of Speech and Hearing Disorders; 28 : 171-182.

Patacek, P.H., Sader, E.K., Maloney, W.and Jackson, C. (1966) : "Phonatory and related changes with advanced age". Journalof Speech and Hearing Research, 9 : 353-360.

Perkins, W.H. (1971) : "Speech Pathology - An applied Behavioural Science". The C.V. Mosby Co., St. Louis.

Ramig, L. (1980) : "Acoustic characteristics of voice and selected measures of body physiology". Unpublished Doctoral dissertation. Purdue University.

Raol, R.J. and Mankame, S. S. (1996) : "Artifical neural networks - a brief introduction". Resonance. 47-54.

Rashmi, M. (1985) : "Acoustic aspects of the speech of children". Unpublished Master's dissertation, University of Mysore.

Saito, S. (1977) : "Phonosurgery. Basic study on the mechanisim of phonation and endolarynageal microsurgery". Otological (Tukwoka) : 23 : 171-384.

Samuel, G.(1973) : "A study of fundemental freqeuncy of voice and natural frequency of vocal tract on an Indian population of different age ranges". Unpublished Master's disseration, University of Mysore.

Sawashima, M. (1996) : "Measurement of Maximum phonation time". Journal of Logopedics and Phoniatrics; 7 : 23 - 28.

Sawashima, M. (1968) : "Movements of larynx in articulation of Japanese consonants". Annual Bulletin, (Research Institute of Logopedics and Phoniatrics., University of Tokyo) : 1 : 11-20.

Schmieder, P., and Baken R.J. (1984) : "The Influence of lung volume on the air flow intensity relationship". Journal of Speech and Hearing Research; 27, 430 - 435.

Shantha, Y.S. (1973) : "Establishing and Validating, Isochronal tone stimulation technique". Master's Dissertation, University of Mysore.

Sheela, E.V. (1974) : "A Comparative study of vocal parametrs of trained and untrained singers". Master's Degree Dissertation. University of Mysore.

Shigemori, Y. (1977) : "Some tests related to the use of air during phonation - Clinical Investigations". Otologia Fukuoka, 23(2) : 138-166.

Shipp, T. and Huntington, D. (1965) : "Some Acoustic and perceptual factors in acute - laryngitic hoarseness". Journal of Speech and Hearing Disroders; 30 : 350 - 359.

Smith, W. and Liberman, P. (1964) : "Studies in Pathological Speech Production", Prepared for Data Sciences Laboratory. Hanscon Field, Mass.

Stare, R.E. (1977) : "Clinical Assessment of laryngeal function" : A symposium. Annual Convention of American Speech and Hearing Association, Chicago.

Starkweather, H. (1961) : "Vocal Communication of Personality and Human Feelings". Jr. Communication, 11 : 63-72.

Stoicheff, M.L. (1981) : "Speaking Fundamental Frequency Characteristics of Non-speaking female adults". Journal of Speech and Hearing Research, 24 : 437 - 441.

Sunil, P., Murthy, H., Venkatesh, C. S., and Vijaya, A. (1996) : "Automatic Recognition of EGG waveforms". Scientific paper presented in ISHA Conference, Madras.

Tharmar, S. (1992) : "Acoustic Analysis of Voice in Children and adults". Unpublished Master's Disseration, AIISH, University of Mysore.

Thurman, W.L. (1958) : "Frequency, intensity, relationship and optimum pitch". Journal of Speech and Hearing Research;, Vol. 1, No. 1, 177-189.

Trullinger, R.W. Emanuel, F.W, Skenes, L.L. and Malpass, J.C. (1988) : "Spectral noise level measurments used to track voice improvement in one patient", Journal of Communciation Disorders; 21: 447 - 457.

Vanaja, C.S. (1986) : "Acoustic Parameters of normal voice". Unpublished Master's Disseration, University of Mysore.

Van Riper, C. (1959) : "Speech Correction : Principles and Methods". Englewood Cliff. N.J. Prentic Hall Inc.

Van Riper, C. and Irwin, J.V. (1958) : "Voice and Articulation", Prentice Hall Inc., New Jersey. (Englewood cliffs).

Venkatesh, C.,S.: Cited in Rashmi, M. "Acoustic aspects of the Speech in children". Unpublished Masters disseration, University of Mysore, 1985.

Verma, S. S., Nandakishore, Raman, C.V., Lakhera, S.C., and Dass, S.K. (1983) : "Prediction, of Some Ventilatory" norms in Healthy Indian Males, 21-69 years of age.

Vieira, M.N., Mclnnes, F. R. and Jack, M. A. (1997) : "Comparative assessment of electroglotographic and acoustic measurements of Jitter in pathologic voices". Journal of Speech and Hearing Research ; 40 (1) : 170-181.

Vonlenden (H), Moore, P., and Timicke, R (1960) : "Larynyeal vibrations : Measurements of glottal wave". Part-Ill. The pathological larynx. Archives of Oto laryngology. 71 : (16-35).

Von Laden, H. and Koike, Y. (1970) : "Detection of laryngeal disease by computer technique". Archives of otolaryngology, 91 : 3-10.

Wedin, S. and Orgen, J.E. (1982) : "Analysis of the fundamental frequency of the human voice and its frequency distribution before and after a voice training program". Folia Phoniatrica; 34 : 143-149.

Wendhal, R. (1963) : "Laryngeal Analog synthesis of Harsh voice quality". Folia Phoniatrica; 15 : 241-250.

Wendhal, R. W (1996) : "Some parameters of auditory roughness". Folia Phoniatrica, 18 : 26-32.

Wendhal, R.W. (1996) : "Laryngeal analog synthesis of Jitter and shimmer, Auditory parameters of harshness". Folia phoniatrica; 18 : 98-108.

Wentworth, E.T. (1940) : "Survey of methods of improvement of pitch usage in Speech as presented in twenty - five current Speech Texts". M.A. Thesis, State University of IOWA.

West, R., Ansberry, M. and Carr, A. (1957) : "The rehabilitation of Speech". III Edition, Harper and Row, New York.

Whitehead, R.L. and Emanuel, F.W. (1974) : "Some spectrographic and perceputal features of vocal fry, abnormally rough and model register vowel phonations". Journal of Communication disorders, 7 : 305 - 319.

Wilcox, K. (1978) : "Age and Vowel Difference in vocal jetter". Unpublished Master's Thesis, Purdue University.

Wilcox, K. and Horii, Y. "Age and Changes in Vocal Jitter". Journal of Geronotology, 35 : 194-198.

Wyke, B.(1967) : "Recent Advances in the neurology of phonation : Phonatory reflex mechanisms in the larynx". The British Journal of Disorders of Communication; 2 : 2-14.

Yanayihara, N, Koike, Y. and Von Leden, H. (1966) : "Phonation and respiration. Function study in normal subjects". Folia Phoniatrica; 18 ; 323-340.

Yanagihara, N. and Koike, Y. (1967) : "The regulation of Sustained Phonation". Folia Phoniatrica; 19 : 1-18.

Yanayihara, N. and Von Leden, H. (1967) : "Respiration and Phonation the Functional examination of laryngeal disease". Folia Phoniatrica; 19 : 153-166.

Yoshiaka, H., Sawashima, M., Hirose, H., Ushijima, T. and Honda, K. (1977) : "Clinical evaluation of air usage during phonation". Japan Journal of Legopedics and Phoniatrics, 18 : 87-93.

Yoon, M.K., Kakita, Y. and Hirano, M. (1984) : "Sound Spectrographic analysis of the voice patients with glothic carcinomas". Folia Phoniatrica, 36 : 24-30.

Zurada, J.M. (1992) : "Introduction to artificial Neural Systems", Boston : PWS Publishing company.