CLASSIFICATION OF HOARSE VOICE BASED ON ACOUSTIC PARAMETERS.

Register No. M9820

A dissertation submitted as part fulfilment for the Second Year M.Sc. (Speech & Hearing) to University of Mysore,

All India Institute of Speech & Hearing
MYSORE - 570 006
May 2000

Dedicated
to
Mummy & Papa

CERTIFICATE

This is to certify that the Dissertation "Classification of hoarse voice based on acoustic parameters" is a bonafied work done in part fulfilment for the Second year degree of Master of science (Speech & Hearing), of student with Register No. M 9820.

Mysore

Director
All India Institute of Speech & Hearing
Mysore - 570 006

DECLARATION

This dissertation entitled **"Classification of hoarse voice based on acoustic parameters"** 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. M9820

May, 2000

CERTIFICATE

This is to certify that the dissertation entitled "Classification of hoarse voice based on acoustic parameters" has been prepared under my supervision and guidance.

Mysore
Mav 2000

Dr. N.P. NATARAJA
Prof. & Head of the Department
of Speech Sciences
All India Institute of Speech & Hearing,
Mysore - 570 006.

ACKNOWLEDGEMENTS

I am extremely grateful to my 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 this dissertation.

I am thankful to Dr. (Miss) S. Nikam, Director, All India Institute of Speech and Hearing, Mysore, for permitting me to carry out this project.

I am grateful to Dr. S.R.Savithri, Mrs. Sridevi, Ms. Lalitha and Mrs. Rohini, for their timely help and co-operation.

I thank Miss Yashoda for helping me through the neural netwoork programs and other support.

My sincere thanks to all my subjects without whom this study would have not been possible.

I thank all my classmates for their help and moral support.

Aparna, Anita, Prarthana, Vasanthi, Amirtha and Archana, thanks for being such good friends and making my stay at AIISH a memorable one.

Ruchi and Sanjay, Thanks for all your advices, encouragement, support and of course entertainment. You both mean a lot to me. Miss 'N'.

Last but not the least, I thank Spaceage Electronic Typing for its wonderful job in preparation of this project.

CONTENTS

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" [Michael and Wendahl, 1971]. The production of voice depends upon the synchrony or co-ordination between respiratory, phonatory and resonatory systems. Deviation in any of these systems may lead to voice problems.

The production of voice is a complex process 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. Therefore, voice needs to be constantly monitered and in the event of abnormal functioning of voice, an immediate assessment should be undertaken which would lead to the diagnosis and not only identifies the voice disorders but also acts as an indicator for the treatment and management to be followed.

The ultimate aim of studies on normality and abnormality of voice and assessment and diagnosis of the voice disorders is to enforce a procedure which will eventually bring back the voice of an individual to normal or optimum level. There are various means of analysing voice, developed by different workers, to note the factors which are responsible for creating an impression of a particular "voice" [Hirano, 1981; Nataraja, 1979; Rashmi, 1985].

Traditionally, the clinicians use visual inspection of larynx and subjective perceptual evaluation of voice quality to diagnose the laryngeal pathology [Yanagihara, 1967]. Subjective perceptual evaluation have had some degree of success in separating normal and pathological voice.

However, it has its own limitation on test-retest and inter-rater reliability [Yanagihara 1967, Koike 1969].

There are objective methods like Electro-Myography (EMG), stroboscopy, ultrasound, glottography, ultra high speed photography, photoelectric glottography, electroglottography, aerodynamic measurements, acoustic analysis, etc., which measure various aspects of voice. Presently acoustic analysis of voice is gaining more importance. Hirano (1981), states that ". . . . this may be one of the most attractive method of assessing the phonatory function or laryngeal pathology because it is non-invasive and provides objective and quantitative data". Acoustic analysis can be done by using methods such as spectrography, peak picking, inverse filtering, computer based methods and others.

In computer based techniques there are many software programs which are designed to extract different parameters of voice. The software program used in the present study "Dr. Speech-version 4" developed and marketted by Tiger DRS, Inc, acquires, analysis and displays fifteen voice parameters from a single vocalization. Also, the program provides voice quality estimates.

Because of the multidimensional nature of the physical characteristics of voice signals, it is not enough to analyze pathological voices using only a single acoustic parameter (such as jitter, shimmer or NNE). Hence, multidimensional analysis using multiple acoustic parameters is necessary.

As the computers have been put into use in almost all spheres of life, attempts have been made to use them for classifying the disorder, 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.

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.

Aim of the present study :

1. To establish normative data concerning the Indian population.
2. To find out whether it is possible to differentiate between normals and dysphonics using the parameters weighted in differentiating the two groups.
3. To compare degree of voice quality estimated by perceptual evaluation with the severity rating scales given by Dr. Speech.
4. To classify hoarse voice from the normal voice using a neural network.

In the present study Dr. Speech software was used to extract the following paramters:

(1) Habitual fundamental frequency.

(2) Jitter percent

(3) Shimmer percent

(4) Fo tremor

(5) Mean Fo

(6) Standard Deviation Fo

(7) Maximum Fo

(8) Minimum Fo

(9) Normalised noise energy

(lO)Harmonic to noise energy

(1 l)Signal *to* noise ratio

(12)Amplitude tremor

(13)Maximum Phonation time

(14)S/Z ratio

(15)Ratio%

Hypothesis :

(1) There is no significant difference between normals and dysphonics in terms of the parameters studied.

(2) There is no significant difference between males and females in terms of parameters studied.

Implications of the study :

(l)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 subject 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 very 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 speech articulation" [Boone, 1983].

The underlying basis of speech is voice. Voice is in elements of speech that provides the speaker with the vibratory signal upon which speech is carried. It serves as the melody of speech and provides expression of feelings, intent and mood to thoughts.

"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 word" [Greene 1964]. The main function of voice is for normal daily communication. It is also used for other professional purposes by individuals such as singers, actors, Radio/TV artists, lawyers, teachers, sales persons and others. These professionals are in need to use their voice efficiently. The inefficient or abuse of vocal system leads to organic changes in the system. This causes loss of voice or abnormal voice. Voice problem may severely disturb communication with others, resulting in considerable economic, social, and psychological disturbances. The demoralising effect on communication is greater in the case of professional users of voice. In addition to this, the human voice serves as sublinguistic purpose of survival such as ventilating emotions such as anger, grief and affection which are essential to the maintenance of psychologic equilibrium. According to Perkins (1971) there are atleast five kinds of nonlinguistic functions of voice. Voice reveals speaker's identity, health, emotional state, personality and aesthetic orientation. Voice is also a carrier of cannotative communicative content.

Voice reveals sex, age, intelligence, regional and socio-economic origin, education and occupation. The physiological factors of genetic endowment of physical structures, the health of the individual may affect the voice. The health of an individual may be indicated by qualities of voice that portray pain, respiratory diseases or by those that show fitness and well being. Voice gives psychological clues to a person's self image, perception of others and emotional health. Self image such as confidence, shyness, and aggressiveness can be identified by voice quality. Conclusively, it can be inferred that voice is more than a means of communicating verbal message, it serves as a powerful conveyer of personal identity, emotional state, education and social status.

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\ The sound source 'S' is an acoustic disturbance superimposed upon the flow of respiratory air and is caused by a quasi-periodic modulation of the airflow due to the opening and closing movements of the vocal folds.

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.

Cotz (1961) opined that the crucial event essential for voice production is vibration of the vocal folds. It changes DC air stream to AC air stream, converting aerodynamic energy into acoustical energy. From this point of view, the parameters involved in the process of phonation can be divided into three major groups:

1. The parameters which regulate the vibratory pattern of the vocal folds.
2. The parameters which specify the vibratory pattern of the vocal folds.
3. The parameters which specify the nature of sound generated.

Hirano (1981) has further elaborated on this, by stating that, the parameters which regulate the vibratory pattern of the 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 amculatory muscles. The physical factors include the expiratory force, the conditions of the vocal folds and the state of the vocal tract. Thus, a co-ordination between the three systems, the respiratory, the phonatory and the resonatory are essential for the production of voice. Variations in these systems would be reflected in voice produced.

A voice disorder exists when a person's voice quality, pitch and loudness differ from those of similar age, sex, background and geographical back ground [Aronson 1980; Boone 1977, Greene 1972, Moore 1971]. In other words, when the acoustic and aerodynamic properties of voice are so deviant that they draw attention to the speaker's voice, then disorder of voice is considered to be present.

Damage to voice by means of either misuse or abuse of voice or any pathology in laryngeal system can paralyse social interaction to a great extend, resulting in considerable psychological, social and economic imbalance. Therefore, the voice problem must be treated immediately after it is identified. The voice disorders are classified in terms of etiologic (cause), perceptual (acoustic) and kinesiologic (vocal hypo function and vocal hyper function). Voice disorders are grouped according to acoustic perceptual attributes as quality, pitch, loudness and flexibility [Johnson, Brown, Curtis, Edney & Keaster, (1956]

Voice quality is the perception of physical complexity of laryngeal tone modified by cavity resonation. Fairbanks (1960) tried to distill voice quality defects into three categories-harhsness. breathiness and hoarseness. Individual variation more often exists in perceptual judgement of voice quality.

Vocal pitch is the perceptual correlate of fundamental voice frequency. Disorders of pitch refer to abnormally high and low pitch voices. Loudness is the perception of vocal intensity. The voice may be too weak or too loud. Flexibility is the perceptual correlate of frequency, intensity and complexity variations. The normal voice possesses adequate pitch, loudness and quality variability during spontaneous speech to convey more intellectual and emotional meaning. In voice disorders, these fluctuations are either inappropriately flattened or excessive.

hi clinical practice, rarely abnormal voice vary along a single dimension of quality, loudness and pitch. Most of the time, even though one may predominate, the others are usually present in different combination and proportions.

Defining hoarseness is a difficult task, because hoarseness is a psycho acoustic term used in broader sense to mean any abnormal voice quality due to laryngeal pathology. The term hoarseness is being understood differently by different groups. To a lay-man it implies a sudden change in voice quality or an unpleasent voice. Several researchers have defined hoarseness of voic:

According to Baynes (1966) hoarseness is a quality of voice that is rough, grating, harsh, more or less discordant and lower in pitch than normal for the individual. Moore, Silverman and Zimmer (1971) define hoarseness as characterised by noise of a relatively high frequency that is produced by transient or highly unstable variations. Casper et al (1981) considers hoarseness as a deviation in the tonal quality of the voice resulting when the vocal cords vibrate in an aperiodic or haphazard manner. Van Riper and Irwin (1978) describe hoarseness in terms of breathiness and harshness. Seth and

Gruthrie (1935) stated that the hoarseness is a tonal quality produced when the vocal folds vibrate in an aperiodic, irregular or haphazard manner.

In spite of several meanings assigned to hoarseness, the common factor invariably noticed is that hoarseness is a phonatory phenomenon rather than a resonatory phemomenon i.e., it is produced by the laryngeal sound generator. Therefore the assumption is that hoarseness is the result of some sort of abnormal vibration of vocal cords.

Fairbanks (1960) classifies voice quality disorders into (a) harshness (b) breathiness and (c) hoarseness. Though Jensen (1965) questioned the validity and reliability of this classification, still it is used. Hoarseness is a common symptom of many laryngeal disorders and many a times, it is the only and the first symptom to be noticed. Literature reveals that hoarseness is related to a large number of laryngeal disorders as listed below:

CONGENITAL		ACQUIRED	
I.	Web of the larynx	I.	Traumatic
II.	Cysts	II	Inflammatory
	a) Cystic hydroma		a) Acute - Non specific
	b) Dermoid cyst		- Specific
	c) Branchial cyst		b) chronic - Non specific
			- Specific
III.	Tumors	III.	Neoplastic
	a) Lipoma		a) Benign
	b) Fibroma		b) Malignant
	c) Leiomyoma	IV.	Paralytic
	d) Chondroma	V.	Miscellaneous
	e) Haemangioma		

Table - 2: List of laryngeal disorders related hoarseness.

Sederholms et al (1992) showed with the help of factor analysis that hyperfunction, breathiness and roughness are good predictors of hoarseness. Harshness and breathiness are two components of hoarseness. Harshness is perceived due to irregularity of vocal fold vibrations [Coleman, 1960; Wendahl 1963; 1966; Moore 1975] i.e., variations or perturbations in both amplitude and time period from cycle to cycle give the impression of harsheness. Breathiness is perceived by escape of air through partially closed glottis and the resultant turbulance noise reduces the harmonic to noise ratio (HNR). Excessive aperiodicity also generates noise and reduces the prominence of the harmonics, hence reducing the harmonics to noise ratio. Thus, hoarseness is defined as a voice quality which clearly contains noise components and that can be labelled harsh and breathy (i.e., source noise elements plus friction noise); its perceived pitch tends to vary substantially. Common description of this quality are 'noisy', 'harsh', 'wet' [Anders et al, 1956].

The primary or common factor in hoarseness is noise of a relatively high frequency that is produced by transient vibrations. These sounds are combined with other phonatory sounds that are frequently at low pitch as the result of laryngeal disease or any other condition that would lower the frequency of vocal fold vibration. The transient disturbances seem to occur on the surface of the vocal folds, particularly along the glottis, but other laryngeal structures may also contribute to the total effect.

Moore (1971) grouped sources of laryngeal transients into four categories (1) Accumulation of sticky mucus secretion in the larynx. Excessive mucus tends to interfere with normal movements of vocal folds by weighing them unevenly and damping their excrusion through causing them to adhere to each other. (2) Relative flaccidity of one or both vocal folds. The flacidity causes independent vibration, resulting in transient disturbances. (3) Additions to the mass of the folds. Mass causes pitch change, hoarseness by weighting, stiffening and influencing vocal fold's compliance. (4) The destruction of all or part of the vocal folds cause random vibration and transients resulting in hoarseness.

Objective measurements (ultra-high speed photography, synchronstroboscopy, photoglottography, EGG) reveal that there are mainly three phases in a laryngeal wave form: (a) Opening phase, (b) Closing phase and (c) Closed phase, with a definite temporal relationship. The relationship among these functions change as vocal output varies. For example, the phases of the cycle vary with different loudness levels (frequency being constant). The closed phase becomes shorter for louder sounds as compared to softer sounds. Other patterns can be produced by other intensity-frequency combinations. However, the following two conditions are present for any normal phonation, even though variation among the patterns may occur [Moore and Thompson, 1965]. (a) All three phases of vibratory cycle can be seen, (b) The motion of the two cords tends to be relatively synchronous and equal in amplitude.

During sustained normal phonation, the length and amplitude of adjacent cycles are generally similar. However, careful observation of a phonatory sequence may demonstrate small changes in the contour showing frequency and amplitude of the cycles, ie., these parameters are rarely precisely the same among cycles within a series. This has been supported by Scripture (1906), Simon (1927) and Lieberman (1961). Thus, the normal laryngeal vibration provides a basis for analyzing vocal fold motion in abnormal hoarse voice. In this respect, it would be theoretically possible for the vocal folds to move in a number of typical ways in individual cycles or sequences of vibrations. The vocal folds could move within a single cycle in atleast five ways: (a) Absence of glottic closure (b) different amplitudes of movement in each cord (c) lack of movement by one cord and (d) dissimilar movement patterns along the extent of one or both folds [Moore and Thompson, 1965].

In addition to the above mentioned abnormalities, laryngeal vibrations possibilty exist with sequences of vibratory cycles. These would include random and patterned changes in the amplitude or a period for successive glottal openings. These changes could occur simultaneously in both folds or independently in either fold. Thus, the potential complexity of vibratory patterns resulting from cyclic abnormalities and sequential irregularities are almost endless. Accordingly, if hoarseness can be assumed to result from abnormal vocal fold vibration, its origin should be found in one or more of the suggested vibratory patterns. Based on the above mentioned assumption, perceptual and acoustic studies either in isolation or together have been conducted to find out the correlation between these two measures of hoarseness.

The perception of normal human voice quality is a multidimentional function dependent upon the interactions among several acoustic parameters that result from variations in laryngeal or superlaryngeal behaviour during voice production.

Acoustic analysis of pathological voices have provided one of the most attractive methods for assessing vocal function. The measures used in acoustic analysis of voice are convenient, non-invasive, objective, sensitive and quantitative method of studying laryngeal mechanism while producing speech. Studies have been conducted to identify measurable voice features that are correlated with hoarseness and thus effectively predict the degree of hoarseness perceived by listeners. Some of the advantages of these methods are that quantitative data from the correlated measurement could be easily stored or transmitted to those who need to see them. The measurements are repeatable from audio recorded voice samples and now standardized procedures are available which predict effectively the degree of hoarseness.

At present various computer based methods have evolved which are very fast in terms of analysing the voice samples and giving the values of the parameters as such. These methods are being used in clinical and research work because they are time saving and they don't need interpretation on the part of experimenter since the parameters are automatically analyzed and the values are given.

One such computer based software program presently being used for assessment of voice is Dr. Speech developed by Tiger DRS. Inc. The parameters considered in this program are.

1. Habitual fundamental frequency
2. Jitter percent
3. Shimmer percent
4. Fo tremor
5. Mean Fo
6. Standard deviation Fo
7. Maximum Fo
8. Minimum Fo
9. Normalised noise energy
10. Harmonic to noise energy
11. Signal noise ratio
12. Amplitude tremor
13. Maximum phonation time
14. S/z ratio
15. Ratio %

Further review would show the importance of each of the parameter in differentiation of normals and dysphancies and in further classification of dysphonics.

(1) Fundamental Frequency (Fo):

Fundamental frequency is a measure of the rate of quasi-periodic vibration of the vocal folds. Fo is reported in Hz. It indicates the number of cycles per second of the vibratory pattern. Anderson (1961) noted than both quality and loudness of voice are mainly dependent upon the frequency of vibration. Thus, it seems apparent that frequency is an important parameter of voice. The study of fundamental frequency obviously has important clinical implications. There are various objective methods to measure the fundamental frequency of voice. Cooper (1974) used spectrographic analysis, as a clinical tool to compare the Fo in dysphonics before and after vocal rehabilitation. Jayaram (1975) and Shantha (1973) found a significant difference in the habitual frequencies between normals and dysphonics.

Nataraja, Jagadeesh and Kumar (1984) measured fundamental frequency in phonation in thirty normal males and females. The mean Fo values (in Hertz) for males and females respetively were 141.49 and 237.03. Number of studies have been carried out to find normal fundamental frequency in different age groups of the Indian population [Sheela. 1974, Jayaram 1975, Vanaja 1986, Nataraja 1986, Sredevi 1987, Tharmar 1992, Suresh 1991, Sanjay 1991, Krishnan 1932, Prabha 1997, Pradeep 1997].

An number of studies have been done on Indian dysphonic population [Jayaram, 1975; Nataraja, 1986; sanjay 1991; Prabha 1997; Biswajit 1995; David, 1998, Binu, 1998, Rajkumar 1998, Preethi 1998]. Results of the above studies show that a significant difference occurs between the Fo values in normals and dysphonics. Thus it is apparent that the measurement of the fundamental frequency is important in the diagnosis and treatment of voice disorders.

(2) Jitter percentage (Jitt):

Jitter is the cycle to cycle variability of the pitch period or fundamental frequency. Jitter is a measurement of how much a given pitch period differs from the one or several pitch periods that immediately precede or follow it.

Percent jitter is defined as mean jitter in milliseconds divided by the mean periods in milliseconds, multiplied by 100.

$$\text{Percent Jitter} = \frac{\frac{1}{n}\left[\sum_{i=1}^{n=1}(P_i - P_{i+1})\right] x100}{1/n \sum_{i=1}^{n} p_i}$$

Moore and Thompson (1965) found that percent jitter was 4.9% and 1.4% for severally and moderately hoarse voices rspectively. Jacob (1968) found a median jitter of about 0.6% for phonation produced at a comfortable pitch and intensity level. Hollien et al (1973) found 0.5% and 1.1% jitter for 102 Hz and 276 Hz sustained vowel phonations. Results of jitter analysis of normal sustained phonation by young adults indicated that jitter of the order of 0.5 to 1.0% was typical [Hollien, Girard and Coleman, 1977; Horil 1979]. Smith, Weinberg, Feth and Horii (1978) established a range from 5.4 to 14.5% of jitter for esophageal voice. Nataraja and Savithri (1990) reported that a jitter greater than 3% is considered abnormal. Biswajit (1995), David (1998) and Anita (1994) found a significant difference in jitter values between the normal and dysphonic groups. Difference was also seen between the males and females of both the groups.

(3) Percent shimmer:

Shimmer or amplitude perturbation is a measure of cycle-to-cycle fluctuation in wave form amplitude.

Percent shimmer tends to be proportional to the absolute amplitude. Hence a correction is required to make this measure free from absolute amplitudes which makes it necessary to divide this measure by mean peak amplitude of these cycles. This ratio is analogous to jitter ratio or jitter factor. This ratio can be converted into percentage by multiplying it by 100 which is called percent shimmer.

$$\text{Percent shimmer} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n-1}(A_i - A_{i+1})\right] x\ 100}{1/n \sum_{i=1}^{n} A_i}$$

Nataraja and Savithri (1990) reported that the percent shimmer of 3% can be considered normal and aboue 3% as abnormal. A significant difference is seen between normals and dysphonics group and also within the males and females of both the groups [Anita. 1994 and Biswajit 1995].

(4) Fo Tremor frequency (FFTR):

FFTR is described as the frequency of the most intensive low-frequency Fo-modulating component in the specified Fo - tremor analysis range Biswajit (1995) found that a significant difference occurs between normal and dysphonic males for the vowel ju| and sentences but not for |a| and |i|.

Anita (1994) reported that there is no effect of using different samples (phonation of |a|, |i| & |u| and sentence) on the values of Fo tremor in normals and dysphonics, both males and females. She noted that dysphonics males and females had higher mean values of Fo tremor than normal males and females, because of the inability of the dysphonics to maintain a constant pitch and intensity in both phonation and sentence.

Biran (1995) stated that no significant difference occurs in Fo tremor frequency across the age group 5-15 years, in both males and females. He found that a significant difference occurs in mean values of Fo tremor frequency between males and females,

(5) Standard Deviation of Fundamental Frequency (SDFo):

Standard deviation of all extracted period-to-period fundamental frequency is known as SDFo, voice break areas are excluded. Studies done on Indian population show that a significant difference between normals and dysphonics occurs [Biswajit, 1995; Anita, 1994]. Anita (1994) observed that there was no relationship between standard deviation of Fo and the different samples used (ie., phonation of |a|, |i|, |u| and sentence).

(6) Maximum Fundamental Frequency (Max Fo) and Minimum Fundamental Frequency (Min Fo):

Maximum fundamental frequency (Max Fo) is the greatest of all extracted period-to-period fundamental frequency values. Here, also the voice break areas are excluded. Minimum fundamental frequency refers to the lowest of all extracted period-to-period fundamental frequency values.

During speech, using a normal phonatory mechanism, a certain degree of variability in frequency is expected and indeed is necessary. Two limited or too wide variation in frequency is an indication of abnormal functioning of the vocal system. An octave and a half in males and two octaves in females is considered normal frequency range. Table-3 shows the values of frequency range in phonation and speech for normals and dysphonics as given by Nataraja and Savithri (1990).

Frequency Range (in Hz)	Normal		Dysphonics	
	Mean	Range	Mean	Range
Phonation	9.00	1.29	210	117-470
Speech	295	117.427	332	121-496

Table - 3 : Shows the values of frequency range in phonation and speech for normals and dysphonics (Nataraja and Savithri, 1990)

Sheela (1974) found has found that the pitch range was significantly greater in trained singers then in untrained singers. Jayaram (1975) reported that in normal males the frequency ranged from 90 Hz to 510 Hz and it ranged from 30 Hz to 350 Hz in dysphonic males. The females of the normal and dysphonic groups presented 140 Hz to 710 Hz; and 60 Hz to 400 Hz as then-range of frequency range respectively. Jayaram (1975) also reported that as a group, dysphonics, both males and females presented a restricted frequency range as compared to normals.

Biswajit (1995) and Anita (1994) found a significant difference in value of maximum Fo between normals and dysphonics. A significant difference was seen in the mean values of minimum Fo between normals and dysphonics by Anita (1994). However, Biswajit (1995) found that no significant difference between normals and dysphonic males for vowels. Thus, the measure of maximum and minimum Fo gains importance in differential diagnosis of dysphonics .

(7) Normalized Noise Energy (NNE):

Normalized noise energy [Kasuya et al., in 1986, 1993] is a measure of the turbulent noise energy produced during vocalization. The NNE is automatically computed from the voice signals using an adaptive comb filtering method performed in the frequency domain. This measure is obtained by subtracting the harmonic signal energy removed by a comb filter. The remained is the NNE. The NNE is given by the following equation :

$$NNE = 10 * \log \frac{\sum_{x} w(n)^2}{\sum_{n} x(n)^2} + BL$$

Where w(n) and x(n), n = 0, . . . N -1 are respectively an estimated vocal turbulent noise component and an original voice signal within 1-4 KHz frequency range, and BL is a compensating for the amount of noise energy removed by a comb filter.

Experiments with voice samples have shown that NNE is especially effective for detecting the glottic cancers, since the NNE measures primarily the turbulence noise caused by the closing insufficiency of the glottis during the phonation. Thus it is very useful in the detection of these diseases. But NNE is not sensitive to the noise caused by irregular vibratory motion of the vocal folds. Hence, NNE is not an effective measure for those laryngeal conditions which produce hoarseness because of a periodicity of vocal fold movements.

Kasuya (1993) stated that NNE provides an indication of the degree of turbulent noise resulting from air leakage through the glottis. The NNE measure is an indicator of breathy voice quality. Although NNE is not the only acoustic correlate of perceived breathiness, it seems to be a more important concomitant of the laryngeal pathologies that produce the breathy voice then do other acoustic measures of vocal function [Kasuya, 1993]. In general, pathological voices appear to have higher NNE values then normal voices.

It has been concluded that NNE may be more sensitive than HNR in detecting the presence of glottal noise and, therefore, more useful in discriminating pathological voices from normal voices [Huang et al, 1992 b: Kasuya, Ogawa, and Kikuchi, 1986;Kasaya, Zue, and Endo, 1993).

(8) Harmonics to Noise Ratio (HNR):

Harmonics to Noise Ratio was proposed by Yumoto, Gould & Baer (1982). It was defined by them as the ratio of acoustic energy of the stable harmonics to that of noise. This measure takes into account, the Jitter and Shimmer present in the signal, which is one of its advantages, because jitter affects the spectrum of a sustained vowel by reducing the amplitudes of harmonics and introducing noise between them. Titze et al (1987) reported that harmonics to noise ratio includes wave form perturbation along with peak amplitude and period perturbation. The first step in the calculation of HNR using signal averaging technique of Yumoto is to average the individual pitch pulses. Here, the size of the averaging window is determined by the largest pitch pulse in the signal. For periods shorter than this maximum, the interval between the end of the pitch pulse and the end of the averaging window is filled with zeros. If a sufficient number of periods are averaged, a large proportion of the noise is cancelled. The RMS energy of the average pitch pulse is used as the numerator in the harmonics to noise ratio calculation. The amount of periodic energy is estimated by successive subtractions of the average pitch pulse from individual periods of the original vowel. The RMS energy in the noise signal is used as the demoniation in HNR calculation on a decibel scale HNR is defined as

$$20 \log \frac{\text{RMS (Average)}}{\text{RMS (Noise)}}$$

Yumoto et al (1982) reported HNR values ranging from 7.0 to 17 dB for a group of normals and from -15.2 to 9.6 dB for a group of speakers with a variety of laryngeal disorders. So as the degree of hoarseness increases, harmonics to noise ratio decreases. They also found a highly significant agreement (P = 0.849) between HNR calculations and the subjective evaluation of the spectrograms. The HNR proved useful in quantitatively assessing the results of treatment of hoarseness. Subsequent researchers found this index to be superior to the other well established indices of hoarseness. Wolfe. Steinfatt (1987) & Wolfe and Ratusanik (1988) reported that the correlation of severity of hoarseness with spectral noise determined in terms of HNR was higher than that with the jitter and Shimmer values measured.

Pathak (1997) found HNR values in normals males as 26.51 and 27.82 for females. Rajkumar (1998) noted the values to be 24.92 for males and 27.33 for females. David (1998), also found similar results, 25.97 and 27.89 for normal males and normal females respectively. Studies have shown that higher values of HNR occurs in females than males.

David (1998) reported that a significant difference occurs in the mean values of HNR between normals and dysphonics.

This algorithm is based on the assumption that a long stationary interval can be obtained from the sustained vowel production, but it can't always be expected in actual recording situations because the speech signal generally has the tendency to change smoothly in amplitude and pitch over a long interval of the sustained phonation. Thus, this method may detect the smooth changes in the waveform incorrectly as noise components. In addition it should be noted that HNR can't quantify noise in the severally hoarse voice that has no recognizable periodic components. Other demerit is that this is highly sensisitive to errors in pitch period demarcation, and a dependency on jitter perturbation Fo and vowel type was also demonstrated. Hiraoka et al (1984) suggested that a voice spectrum should be resolved into three points - Fo component, the harmonic component and the noise and that the relative increase of Fo component in hoarse voice spectra is important. So relative harmonic intensity (Hr) was proposed to evaluate hoarse voice. Relative harmonic intensity is defined as the intensity of the second and higher harmonics expressed percent of the total voice intensity.

$$Hr. = \frac{pi}{p} x100$$

Where pi = intensity of the I th component

P = total voice spectral intensity.

They reported relative harmonic intensity values of 67-72% larger than critical values, for normals and lesser relative harmonic intensity for hoarseness thus providing a good discrimination between normal and hoarse voice.

The HNR given by Yumoto et al (1982) is considered best suited for the quantification of spectral noise due to the reason that HNR is sensitive to both jitter and additive noise. However, it used the vocal output near the lip as a signal for calculation, which can't be considered as the glottal signal because the transfer function of the supraglottal structures modify the glottal source before it is picked up by the mic near the lips.

For this reason, cepstral analysis was employed by Anantha Padmanabha (1992) to nullify the effects of transfer function of a supraglottal cavity. He found that the HNR calculation based on the cepstral analysis of the lip radiated vowels are even sensitive to detect the severe and profound categories of hoarseness, whereas the techniques of Yumoto fails to do so. He developed a software program called harmonics to noise ratio based on cepstral analysis where he used a haming window of only 4 pitch period at a time at an interval of every 10 ms for cepstral analysis and a final cepstrum was obtained by finding a cepstral average (LTCA). Two different measures, peak harmonics to noise ratio and average harmonics to noise ratio, can be obtained. This technique eliminates another disadvantage of traditional harmonics to noise ratio calculation which was found to be sensitive to the smooth changes in amplitude and pitch in addition to actual shimmer and jitter.

A new method of computation of HNR called the Dynamic Time Warping (DTW) was proposed by Qi.y (1992) to avoid the demerits of the earlier methods. In this method, noise components of voice were calculated from the discrepancies between wavelets after they had been optimally aligned in time. The optimal time normalisations of wavelets was accomplished by DTW. This method was evaluated using both synthetic and natural voices and significant reductions in noise were obtained. HNR measure obtained by this technique was free of frequency perturbations.

Another cepstrum based technique was used by Kekrom (1993) to calculate spectral HNR descreases in speech signals. He found that HNR almost linearly with both increasing noise levels and increasing jitter continuum. He concluded that the method could be considered as a valid technique for determining the amount of spectral noise and as a useful measure in the analysis of voice quality.

The results reported by Pathak (1995) also agreed with the above findings. He reported that Peak HNR measures together obtained from the samples of vowels |a| and |u| had the potential to be included in the diagnostic battery for the classification of various degrees of hoarseness and also as a screening measure.

Studies done on Indian populations has shown that a significant difference exists in HNR between normals and dysphonics and also among the males and females group.

(9) Signal to Noise Ratio (SNR)

Kojima Gould. Lambiase and Isshiki (1980), used Fourier expansion to separate the noise from the periodic components, to compute the signal to noise ratio as an objective estimate of hoarseness. The resoltuion of voice into signal and noise components may not be satisfactory, since only three pitch periods are used in the fourier transform ie., one third of the Fourier components was counted as the signal [Hiraoka et al 1984]. This method has theoritical limitations also with regard to the accuracy of estimated noise levels, since the fourier co-eficients derived from a signal with duration T provides estimates of the noise has a continuous frequency spectrum. This method is too complex and time consuming to apply to clinical use.

(10) Amplitude Tremor Frequency :

Amplitude Tremor Frequency is the frequency of the most intensive low-frequency amplitude modulating component in the specified amplitude tremor analysis range. Anita (1994) and Biswajit (1995) reported that a significant difference is observed in Fatr between the normal and the dysphonic group. Anita (1994) found that no significant difference occurs in the mean values for normals females between |a|, |i| and |u|. However, in normals males significant difference was seen between |a| Vs |u| and |i| Vs |u|. In case of dysphonic males and females it was found that there was no effect of using different samples phonation of |a|,|i|, |u| and on fatr values.

(11) Maximum Phonation Duration (MPD):

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 phonation duration with a large air escape suggests a neuromuscular deficit such as laryngeal nerve paralysis.

'Norms' for maximum phonation duration vary from 10 secs for consonants in children to 30 secs for vowels in adults [Arnod, 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 secs) than for females (15-25 secs). Bless and Saxman (1970) studied MPD in boys and girls aged 8 and 9 years and found the MPD for girls was 19 secs 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 betrween 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, Maloney 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; Cooms, 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 :

1. Total air 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 utilised 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 signers. The phonation duration ranged from 15 to 24 secs in trained singers and in untrained singers it ranged from 10 to 29 secs. Sawashima (1966) found no significant difference in the MPD in standing or sitting positions.

Sanders (1963) found MPD with twelve trails 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 14th 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 influenced 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 intergration 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.

Ptacek 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 increased as a function of increase in intensity, and high frequency phonations tended to be rated as more breathy, than 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 tratment 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 co-ordination, 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), Preethi (1998), Binu (1998) have shown a significant diffrence between normal males and females in the maximum phonation duration. Thus MPD has been formed to be useful in differentiating normal and dysphonics.

(12) S/Z Ratio :

Michel and Wendahl (1971) suggested maximum phonation duration of sustained blowing as a possible aerodynamic measure which provides an estimate of the amount of control of respiratory system and which can hence be used to evaluate the voice and its disorders. It is defined as the maximum length of time an individual can maintain an oral airflow.

According to Boone (1971) the clinical evaluation of vocal fold function should consider not only the maximum phonation time but it should be contrasted with a sustained expiration without phonation. He suggests the ratio between |S| and \Z\. |S| being a voiceless fricative and |Z| being a voiced fricative, to assess the function of respiration and phonatory systems.

Boone (1971) stated that "the typical prepubertal child can sustain the voiceless exahalation for about 10 secs. The dysphonic patient without vocal fold pathology will typically be able to extend the voiceless S-S-S and the voiced Z-Z-Z for about the same length of time". While a shorter than normal maximum phonation duration would indicate difficulty at the level of the larynx, a short maximum phonation duration could also be the result of reduced vital capacity. Thus this measure of S/Z ratio not only reflects the laryngeal function, but also gives information regarding the rspiratory system.

Rashmi (1985) studied the maximum duration of |S| and |Z| in 110 male and 100 female normals age ranging from 4 years to 15 years. The results indicated no significant difference in maximum duration between |S| and |Z|, both in males and females, throughout the age range studied.

In male subjects, the maximum duration for |S| at four years was 10.38 secs, and it did not show change upto 11 years. After 11 years of age a decrease was noticed upto the age of 15 years. The maximum duration was seven secs at 15 years which was 10.38 at the age 11 years. A similar trend has been observed in the case of females also. No significant difference between males and females, was observed, throughout the age range studied.

According to Vanaja (1986) the maximum duration for |S| and \Z\ both in males and females, decreased with age i.e., from a mean of 11.3 secs, at 16-25 years age group to a mean of 7.35 secs, for the group 56-65 years for |S| in the case of males. No significant difference was found between males and females at any age studied. It was also noticed that the S/Z ratio was approximately 1.00 for all the age groups, both in males and females.

Boone (1983) studied the S/Z ratio in three groups of subjects. Group-1 consisted of 28 subjects with vocal modules or polyps and group-2 consisted of 36 subjects (with functional dysphonia) and Group-3 was the control group of normals. The subjects with functional dysphonia with no laryngeal pathology and the normal speaking subjects all sustained |S| and |Z| for about the same duiration. Their subsequent S/Z ratios approximated 1.0. The |S| duration of the subjects with nodules or polyps was the same as the subjects in the other two groups. However, the duration of |Z| in the laryngeal pathology group was markedly reduced. The means of the S, Z ratios of the pathology group was 1.65. Such a large contrast suggests that there is a marked decrement in laryngeal functioning. Rashatter and Hyman (1982) who studied the maximum duration for |S| and :Z| in children with vocal modules, also reported that generally, those with laryngeal pathology showed a shortened maximum duration for |Z|.

Preethi (1998) conducted a study on normals and dysphonics and found that no significant difference occurs between the S/Z ratio of normal males and females. A significant difference in S/Z ratio was observed between normal and dysphonic females, but such a trend was not seen among the normal and dysphonic males. Majority of the studies have shown that the S/Z ratio is capable of differentiating dysphonics from normals.

Quantification of Hoarseness :

A) Perceptual analysis of hoarseness :

Human ears have the ability to identify and recognize speaker's voice. Well trained voice clinician are often able to determine the causative pathologies on the basis of psychoacoustic impression of voice [Hirano, 1975].

Voice quality is a term that subsumes a wide range of possible meanings, conveying both supra laryngeal and laryngeal aspects [Kratt and Kiatt 1990]. As to the concept of normal verses pathological voice qualities, some parameters such as diplophonia or aphonia must be regarded as pathological. For most voice quality parameters, however, there *is* no distinct border between normal and pathological. For instance, some studies have shown that breathy voice seems to be a common female voice [Henton and Bladon 1985; Bless, Biever and Campros 1989; Sodersten and Lindestad 1990], whereas creaky or vocal fry is a normal male voice characteristics [Henton and Bladon 1988].

Voice quality is also to some extent culturally conditioned and will likely be influenced by aspects specific to a certain language community. Thus vocal fry/creaky is more common in certain regional accents such as modified Northern English in the area of Leads. Yorkshire [Henton and Bladon 1988]. However, significant correlation between frequency perturbation and perceptual qualities such as instability, roughness, flutter, diplophonia and creakiness/vocal fry were found. This is in agreement with the findings of Deal and Emanual (1978) and Askenfelt and Hammarberg (1986). Hammarberg and Gauffin (1986) concluded that perceptual evaluation by well-trained listeners is reliable and reproducible and can be used for systematic evaluation purposes, if handled with precaution. These researchers further concluded that voice quality can be more precisely perceived, if professional terminology is given to the listener.

The reliability of perceptual evaluation can be improved by (1) Operationally defining the voice parameter to be evaluated, (2) Dlustrating the voice quality parameters by samples of tape recordings, (3) Searching for acoustic and physiological correlates of perceptual parameters.

i) Importance of perceptual evaluation :

Although voice properties can be examined at the physiological, acoustical and perceptual level, the judgement of voice quality is primarily a perceptual matter. Gauffin and Hammerbeg (1995) quoted that even though perceptual voice ratings are subjective and impressionistic, there are factors which points to the need for such ratings: i.e., (i) perceptual aspects of voice are important, as they play a crucial role in the listeners acceptance of the voice (ii) skilled voice clinicians are able to perceive and distinguish between different voice qualities (iii) perceptual training as a means of patient's self/contol of voice function in voice therapy to make the patient improve vocal behavior (iv) as pointed out by Kreiman et al (1993). Isteners judgements are usually regarded against which other (quantitative) measures are evaluated.

ii) Factors affecting perceptual judgements :

(a) Listener groups : Yumato, Sasaki, Okamura (1984) found correlations ranging from 0.51 to 0.79 when 8 laryngologists rated the hoarseness of 87 voices on a 4 point equally appearing interval scale.

(b) Rating scales : Most of the perceptual studies use rating scales to simplify the procedures. But it has limitations like scale being too small leading to lack of adequate information.

(c) Language : Perceptual analysis is always language specific, semantic contents, idiomatic expressions can vary with different speakers.

(d) Age, Sex, social cultural factors : Influence inter judge agreement according to Sonninen and Sonninen (1976).

(e) Voice sets : There is abundance of evidence that listeners differ systematically in their judgements for different voice sets.

(iii) Recent methods in perceptual evaluation :

Toner, Emanuel and Parker (1990) compared two techniques - Direct magnitude estimation and equal appearing interval along with spectral noise level (SNL) measurement and concluded that a high degree of linearity exists between both the techniques and Spectral Noise Level. Multidimensional scaling analysis has been used by Kreiman, Gerratt and Precoda (1990) and Kreiman, Gerratt; Precoda and Berke (1992). Askenfelt and Hammerberg (1986) found PERC-7 technique and that also a multidimensional scale. Weavers and Lowe (1990) used a visual analogue scale in which different relevant parameters were represented by a 100 cm. continuous line, the extremes of which corresponding to non-existance and extremely high occurrence of the trait, respectively. Here, the judges were supposed to mark a point on the continum which they though was is a representative measure of the parameter in the voice under consideration.

(iv) Reliability of perceptual evaluation :

Kreiman et al (1993) found that reliability of the listeners varied greatly from study to study. Also, their own experiment indicated that the ratings varied widely across individual clinicians. They suggested that the presentation of anchor stimuli or taped reference examples of deviant voice qualities might improve, between listener rating consistency.

(v) Scales of voice quality :

It is obvious that the variety of voice quality dimensions in a voice sample is to be rated should be reflected in the rating parameters and scales. The most prevalent scale has been the Equal Appearing Interval (EAI) scale [Kreiman et al 1993] which requires the listener to assign a number 1 and n (most of 5 or 7) to a voice sample regarding degree of a certain voice quality. The advantage of EAI scale is easier communication because of the numbers. Kreiman et al (1993) study indicated, however, that EAI rating driffed in a consistent direction within a listening session.

Correlation between perceptual and acoustic measures :

Many studies have been done to find out the correlation between the perceptual and acoustic measures. Such studies were done by Hartman and Cramon (1984) and Imaizumi (1986). These studies reveal that there is a good correlation between acoustic parameters studied and amount of hoarseness perceived.

(i) "It is also known that perturbations with large magnitude give rise to rough voice quality" [Wendahl 1963; Coleman, 1971].

(ii) "This connection between perceived roughness and wave form irregularities exists independent of whether the irregularities are caused by amplitude perturbations or frequency perturbations [Wendahl, 1966].

(iii) Investigations of the acosutic waves of synthesized complex sounds and human phonations have revealed that rapid, random variations in the periods and amplitudes of successive cycles are associated with the perceived roughness of the signal [Levely and Emanuel, 1970].

(iv) Excessive amounts of jitter and shimmer have been implicated as an indication of laryngeal dysfunction, however, and also together with spectral noise components, correlates with hoarse voice quality [Heiberger and Horii, 1982].

(v) Askenfelt & Hammerberg (1986) compared the perturbation measure with regard to the acoustic perceptual correlation and their ability to discriminate between normal and pathological voice status and reported that the standard deviation of the distribution of the relative frequency differences was the most useful acoustic measure for clinical applications.

(vi) Huang (1995b) based on an investigation between perceptual judgements and acoustic parameters made the following conclusions :-

(a) The perception of hoarse voice quality should be considered as some combinations of breathiness and harshness.

(b) Vocal jitter appears to be related primarily to harsh voice quality.

(c) The magnitude of glottal noise closely corresponds to the breathy voice quality.

(d) Shimmer appears to be the primary influence on hoarse voice quality.

(e) The spectral tilt of the glottal source is significantly related to the perceived breathiness.

David (1998) compared perceptual evaluation of normal and dysphonic voice samples on a 4 point scale with objective values i.e., scores of different parameters. He found that quality of voice, normal to severe hoarseness occurs on a continuum and the boundaries between the normal and mild mild and moderate and moderate to severe or not very clear, particularly perceptually. However, the classification of normals and different degrees of hoarseness was possible.

Thus it is seen from the review of literature that many researchers have carried out studies concerning various parameters of voice. However, there are no such studies relating these parameters of voice for both normals and dysphonics concerning the Indian population i.e., using Dr-Speech software developed by Tiger DRS, Inc. Hence, the present study was undertaken to :

(1) To develop a normative data for the different parameters utilized in Dr.Speech software.

(2) Comparing the degree of voice quality estimated by perceptual evaluation with the severity rating scales given by Dr. Speech.

(3) To device an objective method for classification of hoarse voice.

Artificial Neural Networks :

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, ANNs 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, 1994J.

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) problem 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 non-linear device. Consequently, a NN, made up of interconnections of neurons, is itself non-linear and is distributed throughout the network. Non-linearity is a highly important property, particularly if the underlying physical mechanism responsible for the generation of an input signal (eg. speech signal) is inherently non-linear.

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 and 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 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 -3 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.

Signal/Information.......> Flow

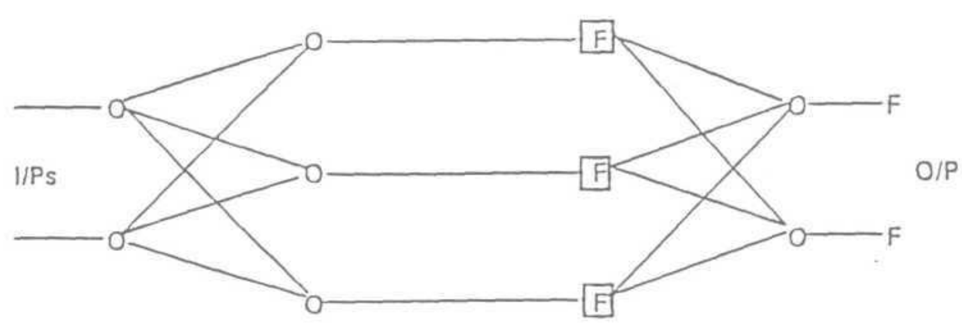


Fig - 3: Typical structure of ANN :

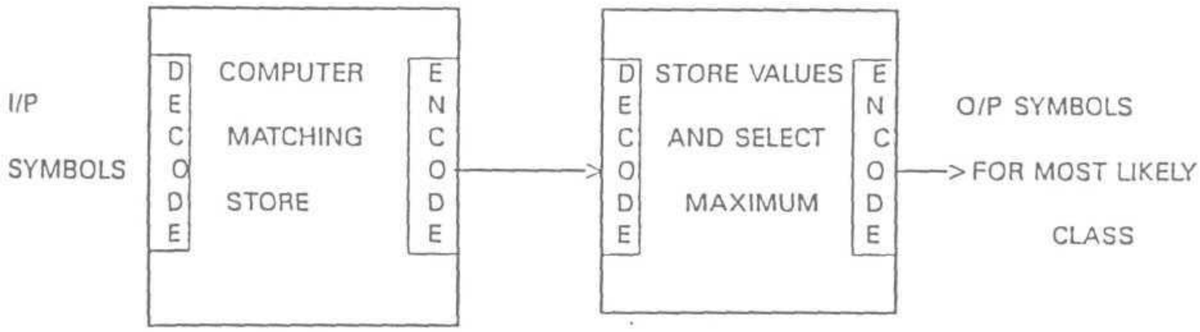
McCullock and Pitts, regarded as pioneers in the field of NNs outlined the first model of an elementary neuron in 1943 [Roal and Mankame, 1996]. However, they could not realise the model using the bulky vacuum tubes of the era. A learning algorithm scheme for updating a neurons connections (weights) was proposed by Donald Hebb in 1949. 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 non-linear 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.

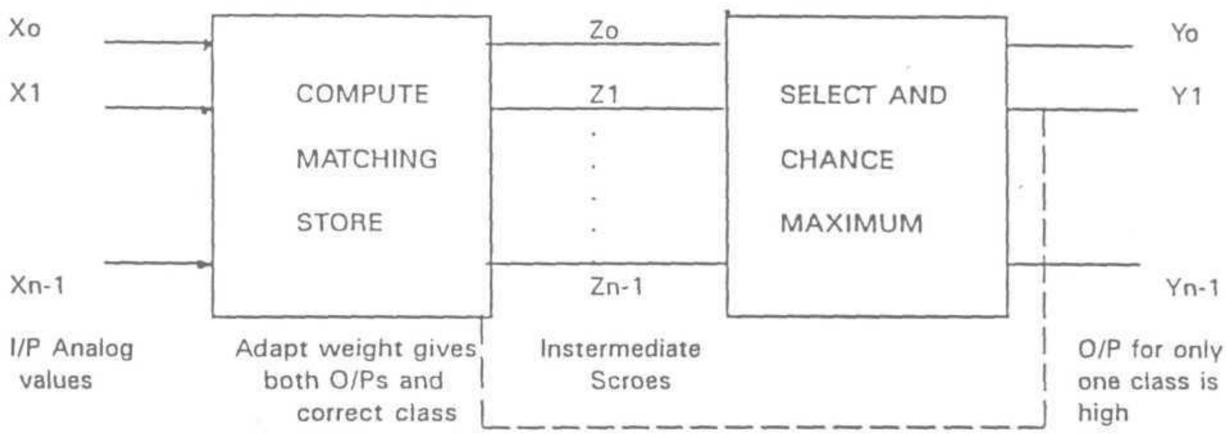
The structure of an ANN is a set of processing units (nodes) arranged in rows. Inputs nodes are interconnected by simple connections with an internal layer of hidden nodes and a single output mode. Rather than having a fixed algorithmic approach to a classification problems, an ANN is sequentially presented 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- 4 as shown.



TRADITIONAL CLASSIFIER (4a)



NUERAL NET CLASSIFIER (4b)

Fig. -4 : Block diagrams of traditional and neural net classifiers.

Both traditional and NN classifiers determine which of M classes is most representative of an unknown static input pattern containing N input elements, hi 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 and 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 elected 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 4. 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 4(a) and 4(b) 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 (finish) 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, Leinomen, Kangas and Torkkola (1993 a) studied word initial samples of fricative (s) preceding vowels [a:], [a2-], [e:], [i:], [u:], [o:] and [y:] in finish 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 mesc intervals, hi all the 10 subjects the [s] samples preceding the round vowels [u:] and [o:] clearly differed from the samples in front of unrounded [a:], [a*.:], [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 [<=£-].

Mujunen. Leinonen, Kangas and Torkkola (1993b) further studied whether the self organizing map also distinguishes beteen 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 (Fo) control by the brain. Good F_o 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-organized 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.

Binu (1998) created an artificial neural network model to classify the voice disorders based on 10 aerodynamic and acoustic parameters for normals Indian population. 30 normal and 414 dysphonic voices were taken as subjects. This neural network developed, was capable of classifying the different types of dysphonics (88.1%). The voice pathologies included were mass on the vocal fold, glottic chink, and congestion. It was also able to differentiate dysphonics from normals.

Hence, from the above review of literature it is evident that various acoustic parameters of voice could be helpful in differentiating dysphonics from normals and further in differentiating different types of dysphonics. 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 data. We know that different parameter of voice show a wide variety of variation in normals and in pathological populations. Hence, the present study aims at creating a model to classify the hoarse voice in terms of its severity i.e., mild, moderate and severe and also to differentiate hoarse voice from the normal voices.

METHODOLOGY

The purpose of the present study was to classify the hoarse voice in terms of various acoustic and aerodynamic parameters using artificial neural networks.

The study was carried out in two phases :

(1) Collection of data on various acoustic and aerodynamic parameters in normals and hoarse voice cases using Dr. Speech software.

(2) Validation of data by comparing the data with other ANN for classification of hoarse voice based on aerodynamic and acoustic parameters.

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

PHASE I: Procedures used to measure different paramters.

The purpose of the Phase I of the study was to examine the relationship between various parameters of voice and hoarse voice. It was decided to establish the normative data of the voice parameters and to differentiate normal and hoarse voice using Dr. Speech software program developed by Tiger, DRS, Inc. for following parameters of voice.

(1) Habitual fundamental frequency.

(2) Jitter percent

(3) Shimmer percent

(4) Fo tremor

(5) Mean Fo

(6) Standard Deviation Fo

(7) Maximum Fo

(8) Minimum Fo

(9) Normalised noise energy

(lO)Harmonic to noise ratio

(1 l)Signal noise ratio

(12)Amplitude Tremor

(13)Maximum phonation time

(14)S/Z ratio

(15)Ratio %

Instrumentation:

The following instruments were used in the present study.

1. Computer Pentium II with 350 Mhz speed
2. Microphone (H legend D800)
3. Sony Tape Deck (TC FX 170)
4. Meltrack audio casette (R - X 90)
5. Philip amp 60
6. Headphones (Sukawa)

Subjects:

30 normals (15 each sex) with an age range of 18 to 25 years (and mean age 20 years) and 3- dysphonics (22 males and 15 females) with an age range of 13 to 68 years mean age 33 years) were studied. The group of dysphonics were chosen from among the patients who visited the All India Institute of Speech and Hearing with the complaint of voice problem. These cases were diagnosed as having voice disorder after the routine speech science, speech pathology, otolaryngalogical and psychological examinations. The normal subjects were taken who had no history of a speech, language or hearing disorders and had under the examination by qualified Speech Pathologists.

SI. No.	Nature of problem	No. of subjects	Age Range (Years)	Sex	Severity of hoarseness		
					Mild	Moderate	Severe
1.	Glottic chink	5	25-34	F	4	1	-
		5	19-66	M	3	-	2
2.	Congestion on vocal folds	16	13-68	M	11	2	3
		11	18-50	F	9	1	1
3.	Mass on vocal folds	5	14-66	M	4	-	1
		5	20-35	F	4		1
4.	Thickening of vocal folds	2	56-59	M	2	-	-
		4	18-42	F		_	1
5.	Vocal fold palsy	1	25	M	1	-	1
6.	Aphonia	1	25	M	1	-	-

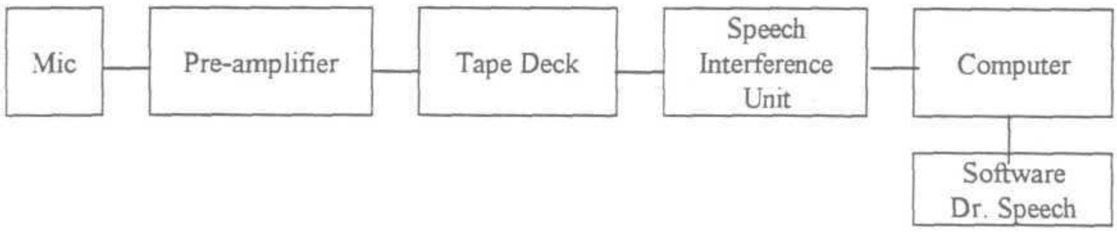
Table -1 : Table showing the distribution of different types of dysphonics both males and females.

Procedure:

(1) Maximum phonation duration (MPD):

MPD has been defined as the duration for which an individual can sustain phonation.

The subject was instructed as follows: "Take a deep breath and say |aj as long as you can. Please try to maintain it at a constant level. Then the subject phonated as long as possible. The duration of |a| was measured using the Dr. Speech program in Speech Science Laboratory at All India Institute of Speech & Hearing, Mysore. Microphone was kept 4 to 6 inches from the subjects mouth. The task was repeated for |i| and |u| also".



Block diagram of the instrumentation set-up

(2) Acoustic Paramters :

These parameters were- simultaneously measured along with measurement of maximum phonation duration for all the vowels |a|, |i|, and|u|. This was done through recorded voice samples. Values for the following acoustic parameters were automatically extracted.

(1) Habitual fundamental frequency.
(2) Jitter percent
(3) Shimmer percent
(4) Fo tremor
(5) Mean Fo
(6) Standard Deviation Fo
(7) Maximum Fo
(8) Minimum Fo
(9) Normalised noise energy
(lO)Harmonic to noise ratio
(1 l)Signal noise ratio
(12)Amplitude Tremor
(13)Ratio%

(3) S/Z **Ratio** :

The S/Z ratio is defined as the ratio of the durations for which the fricative |sj and |z| were produced by the subjects i.e.,

$$\text{S/Z Ratio} = \frac{\text{Maximum duration of sustained" } |s|}{\text{Maximum duration of sustained } |z|}$$

The maximum duration for which the subject could sustain |s| and |z| were determined using the Dr. Speech program.

Live Vs Recorded samples :

For 5 normal subjects, comparison was made between the live and recorded samples. Both the samples were recorded simultaneously. The values obtained on each parameter were compared between the two samples for each subject. 'T' test was used to make this comparison.

Perceptual Analysis :

The phonation and speech samples of both 10 normals and 37 dysphonics were randomly dubbed into another audio cassette. No identity was revealed about the subject on dubbing, except case number. For intrajudge reliability check. 5 abnormal samples were randomly selected and recorded second time.

Seven judges (males and females) from M.Sc. (Speech & Hearing) students were selected for perceptual evaluation. Scoring was done on a 4 point scale.(Normal -0, Mild-1, moderate -2, severe -3) for this purpose, perception lab in speech science department of All India Institute of Speech and Hearing was used. The samples were presented through ear phones at a comfortable loudness.

A comparison was made between the perceptual estimation of hoarse voice with the severity ratings of hoarse voice given by Dr. Speech on a 4 point scale (0 - normal, 1 - mild, 2 - moderate, 3 - severe).

Statistical Analysis:

SPSS (Statistical Package of Social Science) program was used for descriptive and discriminant analysis. Descriptive analysis was done to calculate the mean and the standard deviation and range of parameters. Then the data was treated with T-test to find out the significance of difference of means and standard deviations of all parameters between and within the groups. Further, the data was treated with disciminant analysis for classification of parameters within and across group.

Reliability test:

Five males and five females from the normal group were used to repeat the measures on Dr. Speech program. The values obtained on each paramter were compared with the values of each parameter of the previous measurement for each subject T' test was used to make this comparison.

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 199 MHz processor.
- Microsoft Windows 95
- (16 MB) Ram
- A hard disk with 2 GB space
- A SVGA monitor, with appropriate graphics card
- A mouse

A neural network consists of a number of processing elements, normally arranged in layers. Each processing element is linked to elements in the previous layer by connections that have an adaptable strength or weights. These networks require both inputs and targets.

The neural network option chosen for the present study was 'Multi-layer Perceptron' (MLP). The MLP consists of a number of simple processing elements, arranged in layers. The inputs to each processing element are usually fully connected to the output of the previous layers (Fig. 1).

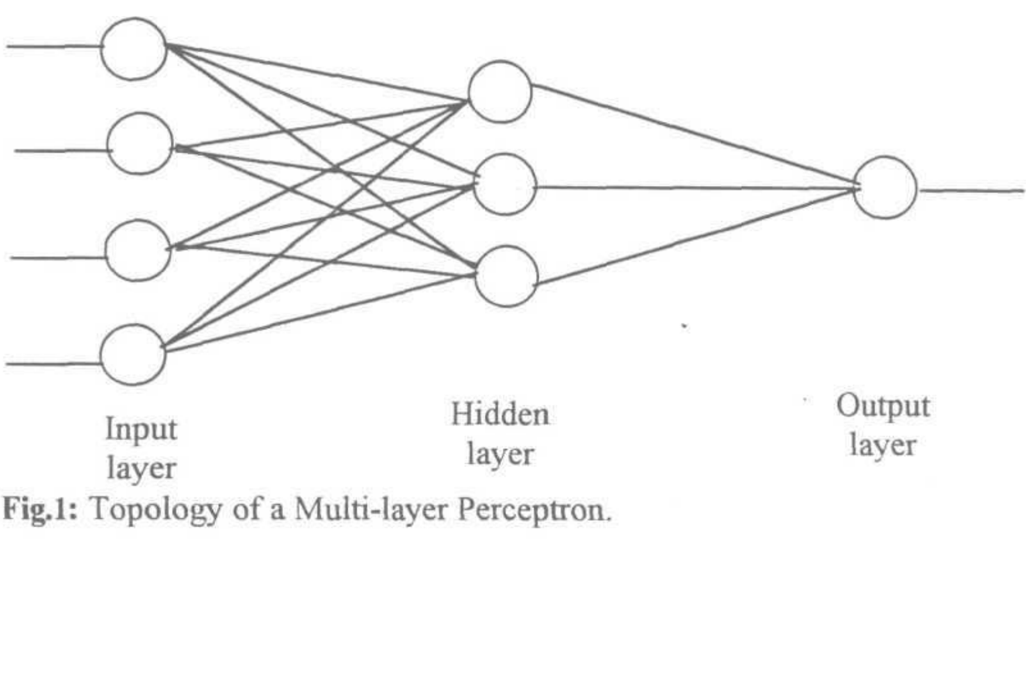


Fig.1: Topology of a Multi-layer Perceptron.

The Figure-2 shows the order of arrangement of the tools which were used for the study.

Fig. 2: Arrangement of tools.

The data of (30 normal and 30 dysphonics subjects) were entered inside the data input tool. There were 15 input fields which included the acoustic parameters. The 16th field was the target field which contained the symbolic representation for normals as 1 and dysphonics as 2. A major portion of the data input tool was allocated for training the data and a smaller percentage of the data was used for validating and testing the data. Hence we have.

Training Data : - It was used to train neural network. Out of a total of 60 data used in the study 48 datas were used for training.

The Multilayer Perceptron was trained by a incremental learning technique where by it was trained in stages using the data of 60 subjects. In the first stage a sample from the training set was used to train the MLP. The best network in this stage was then passed to the 2nd stage and was used as a starting point for training. In 2nd stage, a large sample of data was used to train the network, and best network was passed on to the next stage. This procedure continued for four stages. The samples from the 60 data were reused for the training purpose. The number of samples trained in each stages were :

Stage -1		100
Stage - II		100
Stage - III	-	100
Stage - IV	-	10,000

The training automatically stopped when the MLP found a global minimal error which was equivalent to 0.011 for the training and validation data or it automatically stopped when the number of samples in the 4^{th} stage came to 10,000. Training could be also be stopped by clicking 'STOP' in the MLP performance window.

Validation Data : Validation data are a small proportion of the training data that were not used to build the neural model. They were set aside and used to monitor the performance of the neural system during training. This was achieved by measuring the error on the validation data at frequent intervals during the training cycle. This prevents the neural network from overtraining. Out of total of 60 data used in the study 6 datas were used for validation.

Test data : It was used to measure the performance of a training application. 6 Datas out of total of 60 were used for testing.

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 phase. Thus the data was submitted to the ANN for analysis and the classification of dysphonics and normals was carried out.

RESULTS AND DISCUSSION

The objective of the study was to find out the parameters which can differentiate between normals and dysphonics. Therefore, it was necessary to

(a) Establish normative data for males and females for the following parameters from the analysis of voice using Dr. Speech program developed by Tiger DRS, Inc.

1. Mean fundamental frequency
2. Maximum F_o
3. Minimum F_o
4. Standard Deviation F_o
5. Jitter percent
6. Shimmer percent
7. Amplitude tremor
8. F_o tremor
9. Normalised noise energy
10. Ratio percentage
11. Harmonics to noise energy
12. Signal to noise energy
13. S/Z ratio
14. Maximum phonation time.

a) Compare dysphonics group with respective normal groups (in terms of sex)

b) To find out whether these above parameters differed significantly between the phonation of vowels |a|, |i| and |u|.

These parameters were measured and analysed using the voices of 30 normal and 37 dysphonic Indian subjects.

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

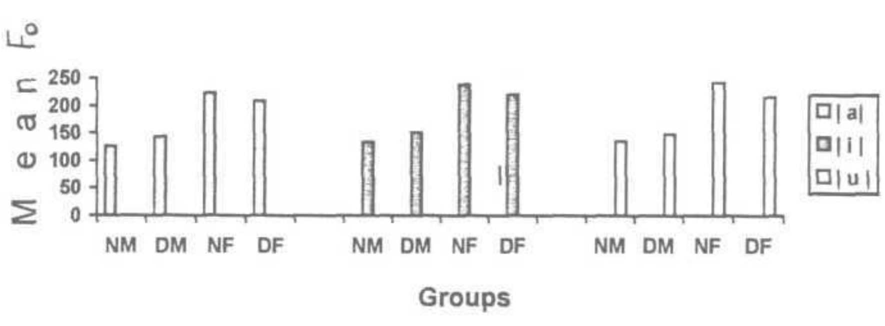
(1) Mean Fundamental Frequency

Mean fundamental frequency was measured using Dr. Speech software. The mean, S.D. and range for mean F_o for normals and dysphonics are presented in Table-4 for the vowels |a|, |i| and |u|.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	125.64	14.12	107.71 - 146.07
	DM	143.2060	43.87	95.32 - 229.67
	NF	223.58	18.39	195.26-259.30
	DF	208.21	26.94	171.99-247.67
|i|	NM	134.15	14.91	112.45- 155.03
	DM	151.84	41.71	98.49 - 250.67
	NF	238.52	17.37	210.27-268.70
	DF	219.91	28.12	168.73 - 256.78
|u|	NM	134.90	13.21	113.59- 158.29
	DM	147.70	42.58	62.11-245.96
	NF	242.11	15.93	215.09-268.75
	DF	215.64	37.12	164.14-270.13

NM = Normal males, DM = Dysphonic males, NF = Normal females, DF Dysphonic females.

Table - 4 : The mean, S.D., range of mean F_o for both normal and dysphonic groups for the vowels |a|, |i| and |u|.



Graph - 1: Comparison of normals vs dysphonics, both males and females in terms of mean Fo in phonation of |a|, |i| and|u|

Vowels	Groups	T values	Significance
|a|	NM Vs NF	16.366	+
	NM Vs DM	-.577	
	DM Vs DF	5.48	+
	DF Vs NF	1.82	-
|i|	NM Vs NF	17.66	+
	NM Vs DM	-1.55	-
	DM Vs DF	5.234	+
	DF Vs NF	2.27	+
|u|	NM Vs NF	20.10	+
	NM Vs DM	-1.14	-
	DM Vs DF	4.73	+
	DF Vs NF	2.63	+

Table - 5 : Comparison of normals Vs Dysphonics, both males and females in terms of mean F_o in phonation of |a|, |i| and |u|.

Table 4 and 5 revealed that there was a significant difference at 0.05 level between males and females of both normal and dysphonic groups for all the vowels.

The males and the females of the dysphonic group showed greater variations than the males and the females of the normal group. Similar findings have been reported by other studies on Indian population [Javaram, 1975; Nataraja, 1986, Rajakumar, 1998].

No significant difference was seen for the vowels |a|, |i| and |u| between normals males and dysphonic males. And, no significant difference between females of both the groups was found for phonation of |a|, However, a significant difference was seen between dysphonic and normal females for |i| and |u|.

Inspection of the Table (4) and T test results revealed that there is no significant difference between |a|, |i| and |u| in the normal male group.

However, in normal females a significant difference was between |a| vs |i| (-2.286) and |a| vs |u| (2.82).

In the present study taking into consideration the mean values and T' test values of mean F_o for the phonation of different vowels, it was found that mean F_o values for normal females was higher compared to dysphonic females. However, in case of males, dysphonic group has higher values of mean F_o than normal group.

The results of this study correlates with the findings of earlier studies.

SI. No.	Investigators and year	Fo in Hz Male	Female
1.	Sheela(1974)	126	217
2.	Jayaram(1975)	123	225
	Nataraja & Jagadish (1984)	141	237
4.	Vanaja(1986)	127	234
5.	Nataraja (1986)	119	223
6.	Sreedevi (1987)	119	218
7.	Tharmar(1991)	124	
8.	Suresh(1991)	131	220
9.	Sanjay(1991)	131	220
10.	Rajashekar(1991)	148	-
11.	Krishnan(1992)	122	231
12.	Pathak(1997)	126	231
13.	Prabha(1997)	125	214
14.	Pradeep(1997)	136	240
15.	Rajkumar(1998)	140	240
16.	David (1998)	127	232

SI. No.	Investigators and year	Male	Female
17.	Preethi(1998)	134	229
18	Binu(1998)	117	230
19	Anita (1994)	129	240
20.	Present study (1999)	125	223

Table - 6: The values of mean fundamental frequency (in Hz) for phonation of normal Indian population as reported by various investigators.

SI. No.	Investigators	Male	Female
1.	Jayaram(1975)	174	202
2.	Nataraja(1986)	152	200
3.	Sanjay(1991)	157	233
4.	Pathak(1997)	141	234
5.	Prabha (1997)	159	217
6.	Rajkumar(1998)	172	229
7.	David (1998)	156	209
8.	Preethi (1998)	154	204
9.	Binu(1998)	152	200
10.	Present Study (1999)	141	208

Table - 7 : The values of mean fundamental frequency (in Hz) for phonation on 'dysphonic' Indian population as reported by various investigators.

Although a difference was seen between dysphonic and normal males groups, this difference was not found to be statistically significant among the groups for the vowel |al, |i| and |u|.

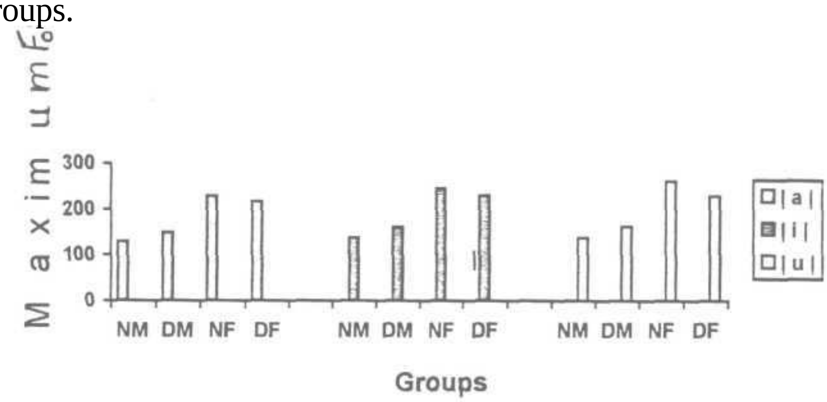
Thus the hypothesis (1) stating that there is no significant difference between normal and dysphonic males in terms of mean Fo was accepted, whereas for normal and dysphonic females it has been rejected. Further the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been rejected.

(2) Maximum Fundamental Frequency

The values of maximum fundamental frequency during phonation for vowels |a|, |i| and |u| for normal and dysphonic groups are presented in the Table - 8.

Vowels	Groups	Mean	**S.D.**	Range
|a|	NM	128.72	13.92	112.21- 149.49
	DM	148.78	43.71	98.44 - 235.83
	NF	228.45	18.88	197.76 - 262.5
	DF	216.36	27.87	177.11-264.91
|i|	NM	137.97	16.64	113.66- 160.95
	DM	161.07	61.67	102.32-308.39
	NF	245.12	19.19	217.24-286.38
	DF	229.69	33.16	175.00-304.14
|u|	NM	138.06	13.94	116.67- 164.55
	DM	162.44	44.04	103.28-280.89
	NF	261.69	18.22	220.60 - 282.69
	DF	228.82	42.69	168.97-304.14

Table -8: The mean, S.D. and range of maximum fundamental frequency in phonation for vowels |a|, |i| and |u| for males and females of both the groups.



Graph - 2: Comparison of normals vs dysphonics, both males and females in terms of maximum Fo in phonation of |a|, |i| and|u|

Vowels	Groups	T values	Significance
|a|	NM Vs NF	16.46	+
	NM Vs DM	-1.712	
	DM Vs DF	5.209	+
	DF Vs NF	1.506	-
|i|	NM Vs NF	16.8	+
	NM Vs DM	-1.664	-
	DM Vs DF	4.372	+
	DF Vs NF	1.559	-
|u|	NM Vs NF	19.18	+
	NM Vs DM	-2.055	-
	DM Vs DF	4.29	+
	DF Vs NF	1.903	-

Table - 9: Comparison of normal Vs Dysphonics, both males and females in terms of maximum fundamental frequency in phonation of vowels |a|, |i| and |u|.

A significant difference was found between the males and females of both normal and dysphonic group for all the three vowels. No difference was seen between dysphonic males and normal males for the three vowels, similarly, even between normal females and dysphonic females no difference was observed.

Anita (1994) found a significant difference in value of maximum Fo between normals and dysphonics. Biswajit (1995) also found a significant difference between normals and dysphonics except for between normal and dysphonic males.

Thus, this parameters did not differentiate the dysphonics from the normals both in case of males and females.

The hypothesis (1) stating that there is no significant difference between normal and dysphonic males and female groups in terms of maximum

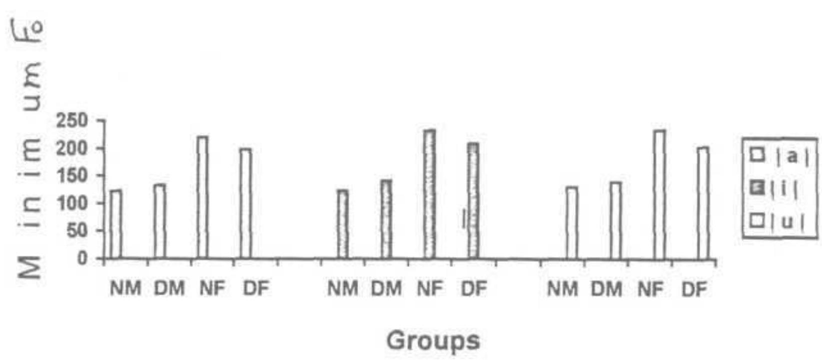
Fo was accepted. Further the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been rejected.

Minimum Fundamental Frequency (Min Fo)

Minimum fundamental frequency refers to the lowest of all exacted period-to-period fundamental frequency values.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	122.16	13.17	105 - 140
	DM	132.61	42.26	60.66-219.40
	NF	219.67	18.83	190.09 - 256.40
	DF	198.77	27.93	162.73-242.31
|i|	NM	124.30	25.88	113.65- 148.99
	DM	142.16	30.66	95.66 - 194.27
	NF	233.07	17.88	206.12-264.07
	DF	208.68	30.35	165.79 - 260.57
|u|	NM	131.30	12.47	110.80- 153.66
	DM	140.33	38.99	60.33 - 225
	NF	234.07	16.93	207.17-264.07
	DF	202.98	36.13	149.49-247.75

Table-10: The mean, the SD and the range of minimum fundamental frequency in phonation of vowels |a|, |i|, and |u|, in both normals and dysphonics groups.



Graph - 3: Comparison of normals vs dysphonics, both males and females in terms of minimum Fo in phonation of |a|, |i| andiui

The males and females of the dysphonic group showed greater variations than the males and females of the normal group. Inspection of Table - 10 and Graph - 3 shows that dysphonic males had a higher mean values compared to normal males, whereas dysphonic females had lower mean values compared to normal females.

The measurement of this parameter in both the males and the females of the two groups, normals and dysphonic are given in Table - 10 and 11.

Vowels	Groups	T values	Significance
|a|	NM Vs NF	16.436	+
	NMVsDM	.915	
	DM Vs DF	5.32	+
	DF Vs NF	2.40	+
|i|	NMVsNF	13.39	+
	NM Vs DM	-1.747	-
	DM Vs DF	6.066	+
	DF Vs NF	2.68	+
|u|	NM Vs NF	18.93	+
	NMVsDM	-.859	-
	DM Vs DF	4.654	
	DF Vs NF	3.00	

Table- 11 : Normals and dysphonics, both males and females in terms of minimum fundamental frequency for vowels |a|, |i| and |u|.

A significant difference was seen between the normal males and normal females and between the dysphonic males and dysphonic females for all the vowels. No significant difference was found between the normal and dysphonic males, however, in case of normal and dysphonic females significant difference was noted. More variations were observed in the dysphonic group than in the normal groups.

Anita (1994) found a significant difference in the mean values of minimum Fo between normals and dysphonics. Thus, the hypothesis (1) stating that there is no significant difference between normal and dvsphonic males in terms of minimum Fo was accepted, whereas for normal and dvsphonic females it has been rejected. Further, the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been rejected.

Thus, this parameter was found useful in differentiating normal and dysphonic female group but not the male groups.

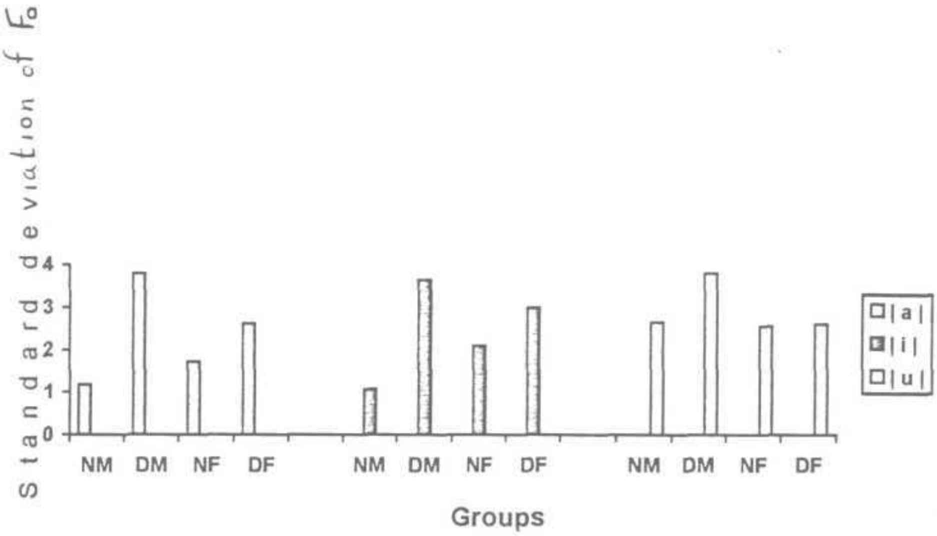
(4) Standard Deviation of Fundamental Frequency (SD F_o):

Standard deviation all extracted period-to-period fundamental frequency is known as SDFo, voice break areas are excluded.

The results obtained in the present study in case of normal and the dysphonic groups (males and females) are presented in Table -12 and Graph -4 as well.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	1.19	.42	.63-2.08
	DM	3.81	9.55	.68-46.30
	NF	1.72	.68	1.12-3.73
	DF	2.62	1.42	.78 - 6.93
|i|	NM	1.08	.46	.46 - 2.03
	DM	3.66	6.72	.86 - 24.24
	NF	2.10	1.33	1.08-6.00
	DF	2.99	1.79	1.09-6.96
|u|	NM	2.65	6.38	.67 - 22
	DM	3.80	9.55	.68-46.3
	NF	2.56	1.73	1.12-7.02
	DF	2.62	1.42	.78 - 6.93

Table -12 The mean, S.D. and range of standard deviation of fundamental frequency for different vowels phonated by both the groups.



Graph - 4: Comparison of normals vs dysphonics, both males and females in terms of standard deviation of Fo in phonation of |a|, |i| and|u|

In dysphonic group no significant difference was found between the males and females. The males and females of the normal groups had significant difference in SDF_0 value were as for the vowel |u| no, such significant difference was seen.

A statistically significant difference was not seen between the normal and the dysphonic group for all the vowels. Also, no difference was seen in SDF_0 for normals males and females between the vowels |a|, |i| and |u|. Amita (1994), also found no difference between the values of standard deviation of fundamental frequency for the vowels |a|, |i| and |u|. Amita (1994) and Biswajit (1995) had reported a significant difference in SDFo between normals and dysphonics group but such a difference was not seen in the present study. Thus, SDFo was not found as a useful parameter in differentiating normals and dysphonic groups.

Thus, the hypothesis (1) stating that there is no significant difference between normal and dysphonics for both males and females group has been accepted. Further, the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

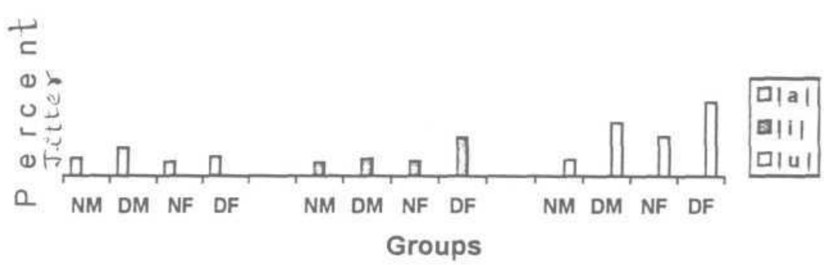
(5) Percent Jitter:

Jitter is the cycle-to-cycle variability of the pitch period or fundamental frequency. Percent jitter is defined as mean jitter in milliseconds divided by the mean periods in milliseconds, multiplied by 100.

The mean, S.D. and range for both the normals and dysphonics are presented in Table - 13 for the different vowels |a|, |i| and |u|. The values of "T' test are given in Table -14 and Graph -5 shows the mean range for the subjects of the two groups.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	.28	.12	.11-.55
	DM	.44	.45	.11-2.25
	NF	.22	.10	.13-.35
	DF	.30	.18	.03 - .76
|i|	NM	.207	.10	.12-.48
	DM	.27	.15	.16-.33
	NF	.23	.13	.10-.45
	DF	.6	.58	.13-2.21
|u|	NM	.25	.16	.12-.75
	DM	.8344	.615	.14 -2.32
	NF	.62	.45	.14- 1.74
	DF	1.14	.7223	.29 - 2.69

Table -13 The mean, SD and range of percent jitter for both normals and dysphonics for the vowels |a|, |i| and **|u|.**



Graph - 5: Comparison of normals vs dysphonics, both males and females in terms of percent jitter in phonation of |a|, |i| and |u|.

The results of percent jitter in males and females of both the groups reveal that no significant differences were seen for the vowel |a| between normal males and females and as well as between dysphonic males and females. A significant difference was observed between the dysphonic females and normal females group in terms of this parameter.

Vowels	Groups	T values	Significance
|a|	NM VsNF	-1.603	-
	NMVsDM	-1.415	-
	DM Vs DF	-1.129	-
	DF Vs NF	-2.3	+
|i|	NM Vs NF	.570	-
	NM Vs DM	-1.187	-
	DM Vs DF	-1.044	-
	DF Vs NF	-1.8	-
|u|	NM Vs NF	2.219	+
	NM Vs DM	-2.143	+
	DM Vs DF	-1.751	-
	DF Vs NF	-2.803	

Table - 14 : Comparision of normals and dysphonics in terms of percent jitter for the vowels |ai, |i| and |u|.

Difference was seen between normal male and dysphonic male group also but it was not statistically significant for the vowel |a|. For the vowel |i| no significant difference was observed between any of the groups. A significant difference was seen for the vowel |u| between normal females and dysphonic females and as well as between normal males and dysphonic males. Also, statistically significant difference was observed between normal males and females, but this was not found between dysphonic males and females groups Biswajit (1995), Amita (1994) and David (1998), found a significant difference in jitter values between normal and dysphonic groups. But this was seen in the present study for the vowel |u| only. Thus, this parameter was not found to be useful in differentiating between normal and dysphonic.

Thus, the hypothesis (1) stating that there is no significant difference between normal and dysphonics for both males and females group int erms of percent jitter has been accepted. Further, the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

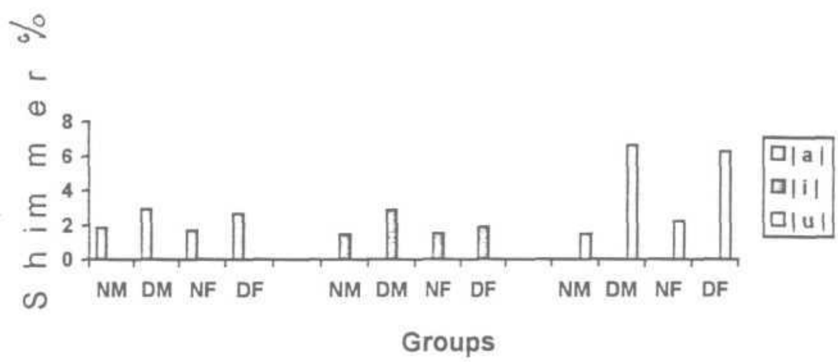
(6) Shimmer Percent:

Shimmer or amplitude perturbation is a measure of cycle-to-cycle fluctuation in wave form amplitude.

The mean, S.D. and range of shimmer percent are presented in table -15 and Graph -6 for both normal and dysphonic, male and female groups.

Vowels	Groups	Mean	**S.D.**	Range
|a|	NM	1.85	.73	1.10-3.45
	DM	2.89	2.00	.96 - 9.06
	NF	1.67	.45	.97 - 2.46
	DF	2.63	1.70	.44 - 7.62
|i|	NM	1.45	.62	.68-2.88
	DM	2.86	2.08	.87-8.61
	NF	1.49	.41	.85 - 2.23
	DF	1.88	.92	.82 - 4.04
|u|	NM	1.46	.49	.73 - 2.28
	DM	6.61	3.16	1.10- 11.91
	NF	2.19	2.03	.82 - 8.08
	DF	6.23	2.77	1.21- 10

Table -15 : The mean, S.D. and range of shimmer percent in phonation of vowels |a|, |i| and |u| in normals and dysphonics.



Graph - 6: Comparison of normals vs dysphonics, both males and females in terms of shimmer percent in phonation of |a|, |i| and|u|

It is evident from the table -14 and Graph - 6 and 'T' test results that there was no significant difference at 0.05 level between normal males and normals females. Similar resutls were seen even between dysphonic males and females.

Vowels	Groups	T values	Significance
|a|	NM Vs NF	-.838	-
	NM Vs DM	-1.919	-
	DM Vs DF	-.570	-
	DF Vs NF	-1.977	-
|i|	NM Vs NF	-2.00	-
	NM Vs DM	-2.524	+
	DM Vs DF	-1.675	-
	DF Vs NF	-1.502	-
|u|	NM Vs NF	1.359	-
	NM Vs DM	-5.033	+
	DM Vs DF	-.355	-
	DF Vs NF	-3.391	+

Table -16 : Comparison of normals and dysphonics in terms of Shimmer percent for the vowels |a|, |i| and |u|-.

A significant difference was observed between dysphonic males and normal males and also between dysphonic females and normal females for the vowel |u|. For the vowel |i|, a significant difference was noted between normal males and dysphonic males but not between dysphonic and normal females. No significant difference was seen for the vowel |a| between normal and dysphonic male and female groups.

Comparison made across the vowels aj, |i| and |u| and 'T' test reveals that there is no significant difference in percent shimmer between the three vowels for the normal group. It is observed from the table that dysphonics had a slightly higher values of percent shimmer than normals and these values were especially much higher in dysphonics for the vowel u| compared to |a| and |i| vowels. Amita (1994) and Biswajit (1995) found a significant difference between normals and dysphonics group and also within the males and females of both the groups. But, such findings were not observed in the present study. Difference was observed between normal and dysphomc males and for the vowels !i| and |u| only and between normal and dysphonic females for the vowel |u|. Thus, shimmer percent is not a very useful parameter in differentiating normals and dysphonics.

Thus, the hypothesis (1) stating that there is no significant difference between normal and dysphonics for both males and females group in terms of percent shimmer has been accepted. Further, the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

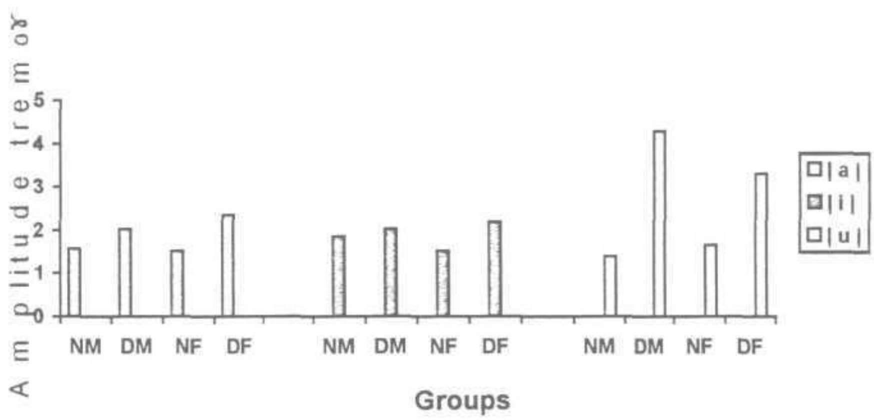
(7) **Amplitude Tremor frequency** :

Amplitude tremor frequency is the frequency of the most intensive low-frequency amplitude modulating component in the specified amplitude tremor analysis range.

Table- 17 and Graph - 7 present the results obtained on this parameter for both the groups.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	1.57	1.71	1-3.42
	DM	2.03	1.76	1-8.58
	NF	1.51	.58	1 - 2.83
	DF	2.35	2.44	1- 10.01
|i|	NM	1.85	1.66	1-7.01
	DM	2.03	1.76	1-8.58
	NF	1.51	.58	1 - 2.83
	DF	2.19	3.22	1 - 13.72
|u|	NM	1.39	.58	1 - 3.04
	DM	4.27	4.42	1.01 - 12.89
	NF	1.64	.60	1 - 2.49
	DF	3.29	3.23	1.03- 11.21

Table - 17 : The mean, S.D and the range of amplitude tremor in phonation of vowels |a|, |i|,]u| for both the groups.



Graph - 7: Comparison of normals vs dysphonics, both males and females in terms of amplitude tremor frequency in phonation of |a|, i| and |u|

While comparing the mean values of amplitude tremor for the phonation of different vowels , it was found that mean amplitude tremor values were higher for dysphonics than for normals. Also, lot of variations were observed in both of male and female dysphonics compared to normals. But, no statistically significant difference was observed among normals and dysphonic groups. Amita (1994) and Biswajit (1995) had reported a significant difference in amplitude tremor between normal and dysphonic group. But, such results were not found in present study. In the present it was found that there was no effect of different samples [phonation of |a|, |i|, and ui] on amplitude tremor. Similar findings were found by Amita (1994). Thus, amplitude tremor was not found to be useful parameter in differentiating normals and dysphonics group.

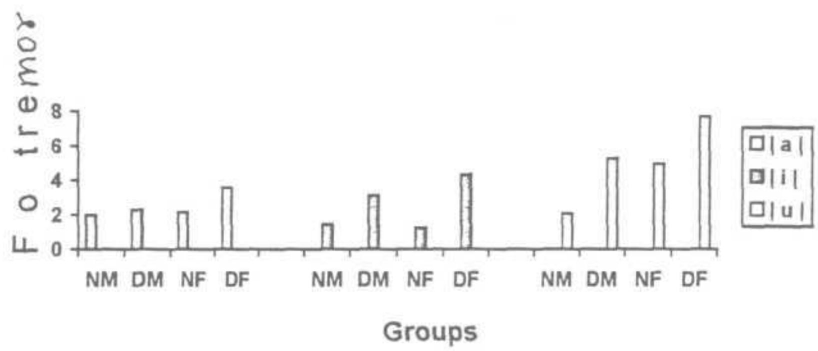
Thus the hypothesis (1) stating that there is no significant difference between normal and dysphonic for both males and female groups in terms of amplitude tremor frequency has been accepted. Further, the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

(8) Fundamental Frequency Tremor (Fo Tremor):

Fo tremor is the frequency of the most intensive low frequency Fo modulating component in the specified Fo tremor analysis range. The mean SD and range for normals and dysphonic for Fo tremor are presented in Table - 18 and Graph - 8. Table - 19 shows the result of the test of significance.

Vowels	Groups	Mean	S.D.	Range
|a|	NM (Normal males)	1.99	1.33	1 - 4.72
	DM (Dysphonic males)	2.28	1.33	1 - 4.89
	NF (Normal Females)	2.12	1.93	1-7.2
	DF (Dysphonic Females	3.56	3.36	1.34. 14.64
|i|	NM	1.44	.56	1-3.10
	DM	3.12	3.43	1.02 - 12.86
	NF	1.24	.26	1-2.01
	DF	4.32	4.35	1 - 13.72
|u|	NM	2.06	1.10	1.01 -4.07
	DM	5.23	4.88	1 - 14.35
	NF	4.92	4.9	1 - 14.88
	DF	7.68	5.12	1.06- 14.92

Table - 18 : The mean, SD and range of Fo tremor in both normals and dysphonics for the vowels |a|, |i| and |u|.



Graph - 8: Comparison of normals vs dysphonics, both males and females in terms of fundamental frequency tremor in phonation of |a|, |i| and |u|

No significant difference was observed between the normals and dysphonic groups for the vowel |a|, |i|, |u|. Also no difference was seen between normal males and normal females and between dysphonic males and dysphonic females for the vowels |a|, |i| & |u.

Vowels	Groups	T values	Significance
|a|	NM Vs NF	-2.50	-
	NMVsDM	-.961	-
	DM Vs DF	.459	-
	DF Vs NF	-.588	-
|i|	NM Vs NF	1.625	-
	NM Vs DM	-1.86	-
	DM Vs DF	.862	-
	DF Vs NF	-2.732	-
|u|	NM Vs NF	2.203	-
	NM Vs DM	-2.464	-
	DM Vs DF	1.380	-
	DF Vs NF	-1.484	-

Table - 19 : Showing the results of test of significance of difference between different groups.

Thus, higher values were obtained for the vowel |u|, followed by |i| and then |a| for Fo tremor. This parameter did not differentiate the dysphonics from the normals for both males and females groups. Also, no difference was observed between males and females of the both normals and dysphonic females. Thus overall this parameter is not useful for differentiating between normals and dysphonic groups and also between males and females.

Thus, the hypothesis (1) stating that there is no significant difference between normal and dysphonic males and female groups in terms of Fo tremor was accepted. Further the hypothesis that there is no significant difference between males and femals for both normal and dysphonic groups has been accepted.

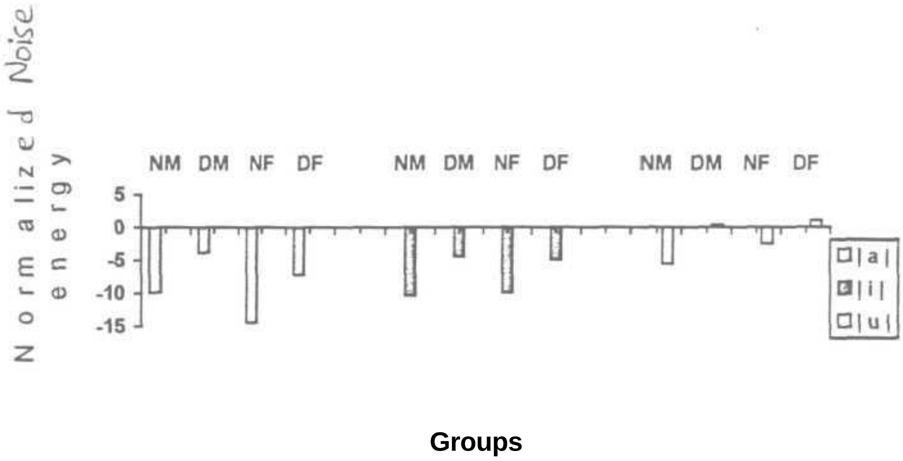
(9) Normalised Noise Energy :

Normalised noise energy is a measure of the turbulent noise energy produced during vocalization.

The values obtained in terms of noise energy for the two groups are given below in Table - 20.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	-9.87	3.42	-15.65-(-45.8)
	DM	-3.87	3.90	-13.67-2.95
	NF	-14.42	2.56	-18.36-(-9.22)
	DF	-7.18	6.25	- 17.72-(-.88)
|i|	NM	-10.41	3.78	-17.16-(-3.71)
	DM	-4.55	3.26	-11.35-(-.19)
	NF	-10.02	3.81	-16.89 - (-3.84)
	DF	-4.99	4.32	-14.82-2.56
|u|	NM	-5.55	3.49	-14.29-(-1.94)
	DM	.43	3.92	-11.85-5.14
	NF	-2.48	2.06	- 7.03 - (-.45)
	DF	1.09	3.36	-4.32 - 7.08

Table - 20 : The mean. SD and range obtained for NNE in normals and dysphonics for different vowels |a|, |i| and |u|.



Graph - 9: Comparison of normals vs dysphonics, both males and females in terms of Normalized noise energy in phonation of |a|, |i| and |u|.

The examination of Table - 20 and 21, Graph - 9 showed that the significant difference occured between the normal and dysphonic groups for both male and female groups in all the three vowels studied. A significant difference was observed between the normal male and female groups for the vowels |a| and |u|. Similar finding was not found for the vowel |ij. The results indicated that a significant difference occured between dysphonic males and females for the vowel al but not for other vowels |i| and |u|.

Vowels	Groups	T values	Significance
|a|	NM Vs NF	-3.377	+
	NMVsDM	-2.229	+
	DM Vs DF	- 2.204	+
	DF Vs NF	-4.8	+
|i|	NM Vs NF	0.350	-
	NM Vs DM	-4.63	+
	DM Vs DF	-.320	-
	DF Vs NF	-3.273	+
|u|	NM Vs NF	2.93	+
	NM Vs DM	-4.572	+
	DM Vs DF	.503	
	DF Vs NF	3.504	+

Table - 21 : Comparison between normals and dysphonics in terms of NNE for the vowels |a|, |i| and |u|.

A significant difference was seen between |a|, |i| and |u| for normal female group [t = -3.724] for |a| Vs |i|, [t = - 5.211] for |a| Vs |u| and [t = -6.513] for |i| Vs |u|. Such a difference was seen in normal males between |ai and |u| (t = -3.153) and |i| and |u| (t = -3.448) but not between |a| and |I|. Thus, it was found that normalised noise energy is a very useful parameter in differentiating normal and dysphonic group for both males and females.

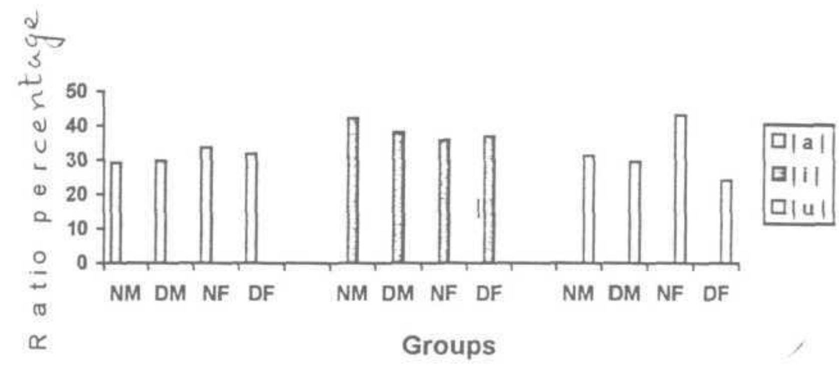
Thus the hypothesis (1) staring that there is no significant difference between normal and dysphonic for both males and females group in terms of Normalised noise energy has been rejected. Further, the hypothesis (2) stating that there is no sigmficant difference between males and females for both normal and dysphonic group has been rejected.

(10) Ratio Percentage:

Ratio percentage is defined as the spectral energy within 2 - 4 kHz ranee.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	29.2	6.47	19-42
	DM	29.88	4.02	22-38
	NF	33.67	6.55	23-47
	DF	31.87	8.53	20-46
|i|	NM	42.2	9.86	25-59
	DM	38.13	8.72	26-48
	NF	35.87	3.98	29-43
	DF	37.07	5.98	24-46
|u|	NM	31.33	6.74	24-41
	DM	29.67	7.48	13-43
	NF	43.13	2.66	38-47
	DF	24.29	10.07	9-46

Table - 22 : The mean, SD and range for the parameter ratio % for both normal and dysphonic, male and female groups.



Graph -10: Comparison of normals vs dysphonics, both males and females in terms of Ratio percentage in phonation of |a|, |ij and |u|

No significant differences were found in the values of ratio between normal and dysphonic groups for vowels |a| and |i|. However a significant difference was observed between normal male and female groups for the **vowel** |u|. A significant difference was also seen between the normal and dysphonic female groups for the vowel |u|, but this was not seen between dysphonic males and normal males.

Vowels	Groups	T values	Significance
|a|	NM Vs NF	1.878	-
	NM Vs DM	-.386	-
	DM Vs DF	.96	-
	DFVsNF	.648	-
|i|	NM Vs NF	-2.307	+
	NM Vs DM	.736	-
	DM Vs DF	-.462	-
	DF Vs NF	-.647	-
|u|	NM Vs NF	7.27	+
	NM Vs DM	.706	-
	DM Vs DF	-.178	-
	DF Vs NF	7.017	+

Table - 23 : Comparision between normals and dysphonics in terms of Ratio Percentage for vowels |a|, |i| and |u|.

Thus, ratio percent is not a useful parameter in differentiating between the normal and dysphonic groups.

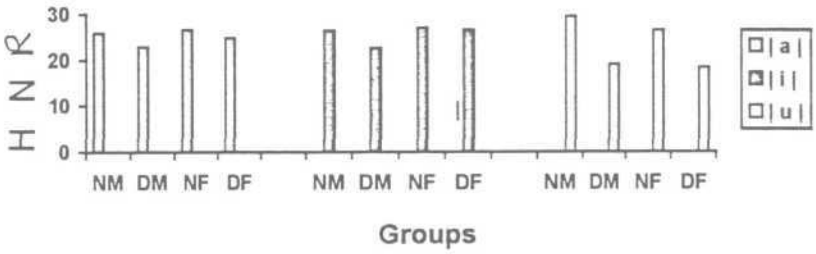
Thus the hypothesis (1) stating that there is no significant difference between normal and dysphonic for both males and females group in terms of ratio percentage has been accepted. Further, the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

(11) Harmonics to Noise Ratio (HNR):

Harmonics to noise ratio is defined as the ratio of acoustic energy of the stable harmonics to that of noise.

Vowels	Groups	Mean	**S.D.**	Range
|a|	NM	25.81	5.12	13.84-25.81
	DM	22.87	7.19	6.32 - 32
	NF	26.53	2.78	22.91-32.21
	DF	24.82	6.85	9.73 - 38.64
|i|	NM	26.48	6.95	3.27 - 32.04
	DM	22.58	6.77	7.07 - 34.9
	NF	26.98	3.86	20.39 - 32.47
	DF	26.50	5.97	15.46-34.36
|u|	NM	29.48	2.32	25.64 -34.48
	DM	18.93	6.63	8.42 - 30.99
	NF	26.42	6.41	13.46-33.48
	DF	18.289	5.83	9.03 - 28.05

Table -24 : Shows the mean, S.D. range for both normal and dysphonics group for phonation of different vowels |a|, |i| and |u|.



Graph - 11 : Comparison of normals vs dysphonics, both males and females in terms of Harmonics to noise ratio in phonation of |a|, |i| and|u|

The results of HNR are shown in Table -24 and Graph - 11. The males and the females of the dysphonic group had no significant difference in terms of HNR. The males and the females of the normal group also showed no significant difference. No statistical difference was found between the dysphonic and the normal groups. Even no statistical significant difference was seen between the phonation of three different vowels among the subjects of normal groups.

The dysphonic group showed lower HNR values than the normal group. The results of this study correlated with earlier studies (Table - 25), (Pathak, 1997; Rajkumar 1998; David 1998).

Investigations	Males	Females
Pathak (1997)	26.51	27.82
Rajkumar (1998)	24.92	27.33
David (1998)	25.97	27.89
Present study	25.81	26.53

Table - 25 : Shows the values of HNR for phonation of |a| on normal Indian population as reported by various investigators.

Thus, HNR was not found as a useful parameter in differentiating normals and dysphonic groups.

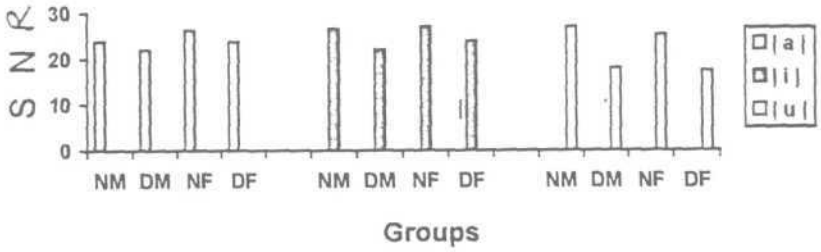
Thus, the hypothesis (1) stating that there is no significant difference between normal and dysphonic for both males and females group has been accepted. Further the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

(12) Signal to Noise Ratio (SNR):

The signal to noise ratio values as observed are presented in table - 26 and Graph - 12. The males and females within the dysphomc group showed no significant difference. This was similar to the results seen within the normal group.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	23.83	6.18	6.26 - 30.52
	DM	21.84	6.96	6.37-31.57
	NF	26.14	2.62	21.75-29.84
	DF	23.69	6.61	9.49-37.13
|i|	NM	26.46	2.99	21.25-31.3
	DM	21.84	6.96	6.37-31.57
	NF	26.84	4.25	19.38-34.40
	DF	23.69	6.61	9.49-37.13
|u|	NM	26.66	6.42	9.11-32.53
	DM	17.82	6.30	8.14-29.84
	NF	24.85	6.41	12.19-32.78
	DF	17.14	5.56	8.57 - 26.69

Table - 26 : Showing the mean, SD and range for SNR for the different vowels in both normal and dysphonic groups.



Graph - 12: Comparison of normals vs dysphonics, both males and females in terms of Signals to Noise Ratio in phonation of |a|, |i| and |u|

The males and females of dysphonic and the normal groups of did not differ from each other for the vowel |a|. For phonation of |i|, a significant difference [t = 3.13] was seen between the dysphonic and normal males, but this did not occur between the dysphonic and normal female groups. A significant difference between dysphonics and normals, both in case of males and females, was observed for |u|.

No sigmficant difference was seen between |\a|, |i| and |u| values of SNR within the normal group i.e., between males and females of normal group. Thus, SNR was not found useful in differentiating between normals and dysphonics group except for the vowel |u|.

Thus, the hypothesis (1) stating that there is no sigmficant difference betwen normal and dysphonic for both males and females group in terms of signal to noise. Ratio has been accepted. Further the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphonic group has been accepted.

(13) S/Z **Ratio** :

S/Z ratio has been defined as the ratio of maximum duration of sustained |S| to maximum duration of sustained |Z| i.e.,

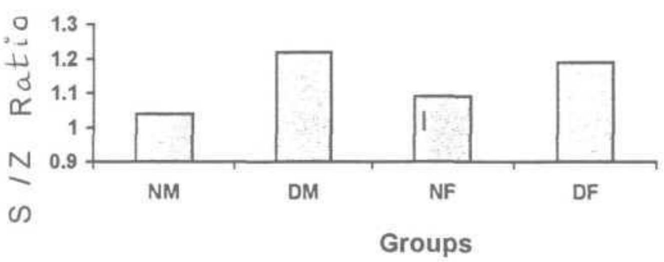
$$\text{S/Z ratio} = \frac{\text{Maximum duration of sustained |S|}}{\text{Maximum duration of sustained | Z|}}$$

Groups	Mean	S.D.	Range
NM	1.04	.16	.8 - 1.33
DM	1.22	.47	.54 - 2.00
NF	1.09	.23	.66 - 1.6
DF	1.19	.35	.72-1.8

Table - 27 : The mean. SD and range of S/Z ratio in both the groups, normals and dysphonics.

Groups	T values	Significance
NM Vs NF	.706	-
NMVsDM	-1.015	-
DM Vs DF	.270	-
DF Vs NF	-.913	-

Table - 28 : Showing the results of the test of significance between different groups for S/Z Ratio



Graph - 13: Comparison of normals vs dysphonics, both males and females in terms of S/Z Ratio.

The normal males showed a mean of 1.04 sec with a SD being 0.16 and normal females showed a mean of 1.09 sec. with SD of 0.23. Inspection of table. - 27 and 28 and Graph - 13 revealed that the S/Z ratio shown by the normal and the dysphonic groups were not statistically significantly different in case of both males and females.Thus, hypothesis stating that there is no significant difference between normals and dysponics is accepted.

Nataraja (1986) and Preethi (1998), found no significant difference between normals and dysphonics in case of males, where as in females a significant difference was seen. In the present study, there was no significant difference observed even in case of females. Thus. S/Z ratio is not a useful parameter in differentiating normals and dysphonics groups.

Thus, the hypothesis (1) stating that there is no significant difference between normal dysphonic for both males and females group has been accepted. Further the hypothesis (2) stating that there is no significant difference between males and females for both normal and dysphomc group has been accepted.

(14) Maximum phonation time :

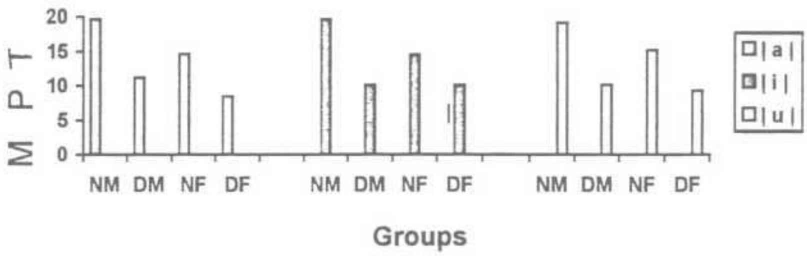
Maximum phonation time (MPT) has been defined as the maximum duration for which an individual can sustain phonation (in sec) after a deep inspiration.

The mean. SD and range of MPT for both the normals and dysphonics are presented in Table - 29 for the vowels |a|, |i| and |u|. The values of 't' test are represented in table - 30 and Graph - 14 show the mean values for the subjects of both the groups.

Vowels	Groups	Mean	S.D.	Range
|a|	NM	19.60	4.81	11-28
	DM	11.19	3.95	6-20
	NF	14.60	4.27	9-22
	DF	8.47	4.03	3- 19
|i|	NM	19.6	4.81	11-28
	DM	10.13	3.59	4- 15
	NF	14.53	3.91	10-21
	DF	10.07	3.53	5 - 16
|u|	NM	19.07	4.22	13-25
	DM	10.11	3.58	5-18

Vowels	Groups	Mean	S.D.	Range
	NF	15/13	4.98	9-27
	DF	9.28	3.89	4- 16

Table - 29 : The mean, SD and range of MPT in both normals and dysphonics for the vowel |a|, |i| and |u.



Graph -14: Comparison of normals vs dysphonics, both males and females in terms of Maximum phonation time in phonafion of |a|, |i| and |u|

Vowels	Groups	T values	Significance
a	NM Vs NF	-3.011	+
	NM Vs DM	6.094	+
	DM Vs DF	.757	-
	DF Vs NF	4.043	+
i	NM Vs NF	-5.677	+
	NM Vs DM	5.604	+
	DM Vs DF	-.046	-
	DF Vs NF	3.283	+
u	NM Vs NF	-3. 001	+
	NM Vs DM	6.630	+
	DM Vs DF	-.626	-
	DF Vs NF	3.504	+

Table-30 : Comparison of normals and dysphonics, both males and females in terms of MPT for the vowels |a|, |i|nd |u|.

It is evident from the study of the Table - 30 that this parameter is able to differentiate between normal males and females for the vowels ai, |i and |u. A significant difference was also found between dysphonic males and normal males and as well as between dysphonic females and normal females. No significant difference was seen between the dysphonic males and dysphonic females group. It was found that both dysphomc males and females had almost the same duration of maximum phonatin, whereas the normal males showed a much longer phonation duration than the normal females. Similar findings have been reported by Nataraja (1986) and Preethi (1998). Thus, maximum phonation duration is found to be very useful parameter for differentiating normal and dysphonics.

Thus, the hypothesis (1) stating that there is no significant difference between normal and dysphonic for both males and females in terms of maximum phonation time has been rejected. Further, the hypothesis (2) stating that there is no significant difference between males and females for normals was rejected, whereas for dysphonics group it has been accepted.

		Significance of difference											
		|a|				lil				|u|			
SI. No.	Parameters	1	2	3	4	1	2		4	**1**	2	**3**	4
1.	Mean Fo	+	_	4-		+		+		+	_	+	+
2.	Max. Fo	+	_	+		+	_			+	_	+	_
3.	Min Fo	+		+	+	+	_	+		+		+	+
4.	SDFO		_	_		_	_	_	_		_	_	_
5.	% jitter	_	_	_	+	_	_	_	_	+		_	+
6.	Shimmer Percent	-	-	-	-	-	+	-	-	-	+	-	-
7.	Amp. tremor	-		-									
8.	Fo tremor			-		_	_	_					
9.	NNE	+	+	+	+		+		+	+	+	_	
10.	Ratio	_	_	_	_	+	_	_	_	+	_	_	
11.	HNR	-	-	-	-	-	-	-	-		-	-	

		Significance of difference											
		|a|				|i|				lul			
SI. No.	Parameters	1	2	3	**4**	1	2	3	4	1	2		4
12.	SNR												
13.	SZ Ratio												
14.	MPT	+	+	-	+	+	+	-	+	+	+	-	_

Table - 31 : Showing the significance of difference on different parameters for the vowels a|, i| and |u|.

1. Normal males Vs Normal females
2. Normal males Vs Dysphonic males.
3 Dysphonic males Vs Dysphonic females.
4. Dysphonic females Vs Normal females.

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. Normalised noise energy.
2. Maixmum phonation time.

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

1. Mean Fo
2. Max. fundamental frequency.
3. Min fundamental frequency.
4. Normalised noise energy.
5. Maximum phonation time.

A significant difference between males and females of the dysphonic group was seen for the following parameters:

1. Mean Fo
2. Maximum fundamental frequency.
3. Minimum fundamental frequency.
4. Normalised noise energy.

Hence, the parameters chosen for the study could sufficiently diffemtiate between normal and dysphonic groups.

Results of Discriminate Analysis :

The discriminant analysis was conducted to examine whether the subjects of the two groups overlapped one another, or diverge from one another in terms of the parameters studied.

The data was processed using the step-wise discriminant analysis with statistical package for social sciences [7.5 windows version], computer program. The criterion for discrimination was taken as 0.05 level. Results of the discriminant analysis of normals and dysphonics are summarized in the Table - 32.

	Group	Predicted group membership 1	2	Total
Original	Count 1	30	0	30
	2	2	28	30
	% 1	100	0	100
	2	6.7	93.3	100

(96.7% of original grouped cases correctly classified)

Table - 32 : Results of discriminant analysis of normals (Group - 1) and dysphonics (group -2) for vowel |a|, |i| and |u| combined.

In discriminant analysis not all the parameters were utilized for the analysis. The parameters used for the classification are following:

1. Maximum phonation time.
2. Normalised noise energy.
3. Habitual fundamental frequency.
4. Maximum fundamental frequency.
5. Minimum fundamental frequency
6. Ratio percentage.
7. Percent jitter.

Thus, the above parameters were found to be sensitive in differentiating the normals from dysphonics.

Comparison of Perceptual and Objective Evaluation :

Perceptual evaluation of the voice (normal and dysphonic) samples rated by 5 judges on a 4 point scale (0 - normal, 1 - mild, 2 - moderate, 3 - severe hoarse voice) was considered. Comparison of the severity rating scales given by Dr. Speech program (0 - normal, 1 - mild, 2 - moderate and 3- severe hoarse voice) was done with the perceptual estimates of voice quality. Pearson's co-efficient of correlation |r| of the judgements of all the ratings for all judges and Dr. Speech were compared. The results of the correlation are summarized below in Table - 33.

Correlations*

		J1	J2	J3	J4	J5	DRSP
Pearson	J1	1.000	.828**	.875**	.511**	.717**	.306*
Correlation	J2	.628**	1.000	.832**	.519**	.800**	.251*
	J3	.875**	.832**	1.000	.593**	.877**	.243*
	J4	.511**	.519**	.593**	1.000	889**	.252*
	J5	.717**	800**	.877**	.669	1.000	.201
	DRSP	.305*	251*	.243*	.252*	.201	1.000
Sig.	J1		.000	.000	.000	.000	.011
(2-tailedO	J2	.000		.000	.000	.000	.039
	J3	.000	.000		.000	.000	0.48
	J4	.000	.000	.000		.000	0.38
	J5	.000	.000	.000	.000		.100
	DRSP	.011	.039	.046	.038	.100	

****** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2 - tailed)

a. Listwise N = 68

[DRSP - Dr. Speech, J - Judges]

Table-33: Showing results of correlation between different judges and Dr. Speech.

The Table - 33 indicates that a positive correlation (.2 to .3) occurs between the perceptual estimates and the ratings given by Dr. Speech. Also, it was observed that the correlation betwen the judges and Dr.Speech values is significant at 0.05 level.

From these results, it was inferred that there was a thin margin between normals and mild degrees of hoarse voice, because many of the mild degree of hoarse voices wee grouped as belonging to normal group. Also, it was noticed that some of the normals were classified/identified as mild hoarse voice cases. This suggests that the quality of voice, normal to severe hoarseness occurs on the same continum and the boundaries between normal and mild, mild and moderate and moderate to severe are not very clear, particularly perceptually. Therefore, overlap cross this boundaries has to excepted/accepted. However, the classification of normals and different degrees of hoarseness has been possible.

PHASE - II:

Validation of data by comparing the data with an Artificial Neural Network for classification of hoarse voice based on acoustic paramters.

Artificial Neural Network

The following fields were taken to be fed into the neural network for processing and classification.

1. Habitual fundamental frequency
2. percent Jitter
3. percent Shimmer
4. Mean fundamental frequency
5. Fundamental frequency tremor
6. Amplitude tremor
7. Normalised noise Energy
8. Signal to Noise Ratio
9. Harmonics to Noise Ratio
10. Ratio Percent

11. Minimum Fundamental Frequency
12. Maximum Fundamental Frequency
13. Maximum Phonation Time
14. S/Z Ratio
15. Group

The data was trained using the multi layer perceptron (MLP) neural networks model by using the data of 60 subjects. Out of the 60 data used in the study 48 data (80%) were used for training, 6 (10% for validation and 6 (10%) for testing.

The test data consisted of 6 samples. The results of the test data are shown in the cross tabulation matrix .

		Predicted	
		b	a
True	b	5	1
	a	0	0

a = Normals

b = Dysphonics

Total number	=	6
Number of correct response	=	5
Percentage of correct responses	=	83.33%

From the Table it can be seen that out of a total of 6 subjects 5 were identified correctly. It was seen that the network was able to correctly identify the dysphonics 83.33% of the times. Better results can be obtained by increasing the number and varieties of cases.

Hence, the neural network proved to be capable of classifying the normals and dysphonics based on the above acoustic parameters.

SUMMARY AND CONCLUSION

The present study was designed to classify hoarse voice based on acoustic parameters. "Dr. Speech" software program was used to acquire, analyse and display the following voice parameters from a single vocalization. These extracted parameters were available as a numberical file which was subjected to statistical analysis.

1. Habitual fundamental frequency (fo)
2. F_o tremor
3. Mean fundamental frequency
4. Maximum F_o
5. Minimum F_o
6. Standard Deviation F_o
7. Jitter percent
8. Shimmer percent
9. Amplitude tremor
10. Normalised noise energy
11. Ratio percentage
12. Harmonics to noise energy
13. Signal to noise energy
14. S/Z ratio
15. Maximum phonation time.

All the 15 parameters were measured in a group of 30 normals (15 males and 15 females) and a group of 37 dysphonics (22 males and 15 females). The results were sybjected to statistical analysis ('T' test and discriminant analysis) using SPSS computer programme.

For perceptual evaluation, the audio recording of speech was used. Five judges were asked to rate the overall severity of hoarse voice on 4 point scale for each sample presented in a random order. Inter and intrajudgements have been found to be reliable.

The input data consisting of 15 acoustic parameters was fed to the Multilayer perception a neural network model to classify normal and dysphonic voice.

The following results were obtained:

(a) It was found that following five variables were more sensitive to differentiate between normals and dysphonics.

1. Normalised noise energy.
2. Maximum phonation time.
3. Mean fundamental frequency.
4. Maximum fundamental frequency.
5. Minimum fundamental frequency.

(b) The discriminant analysis showed that 96.7%, of original grouped cases (normal and dysphonic group) were correctly classified.

(c) No significant differences were observed between most of the parameters for the vowels |a|, |i| and |u.

(d) It was found that a positive correlation (..2 to .3) occurs between the perceptual estimates and the ratings given by Dr. Speech program.

(e) Results of the neural network model revealed that Multilayer perception could correctly classify 83.33% of the cases correctly.

Conclusion:

1. Five out of the fifteen parameters, studied were found useful in differentiating betwen normal and dysphonic groups.
2. The acoustic parameters were capable of classying normal and dysphonic voice.
3. The neural network can be trained using 15 parameters to classify normals and dysphonics.

Recommendation:

1. A similar study may be conducted in larger samples.
2. Differentiation of hoarse voice based on severity may be done using neural network.

BIBLIOGRAPHY

Ananthapadmanabha, T.V. (1992). HNR program based on cepstral analysis. Personal Communication.

Anders, L.C., Hollien, H., Hurme. P., Sonnineiu A. and Wendle, J. (1988). Perception of hoarseness by several classes of listeners. Folia Phoniatrica, 40(2), 91-100.

Anderson, V (1961). Training the Speaking Voice Oxford University. New York.

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

Askenfelt, A.G. and Hammarberg, B. (1986). Speech waveform perturbation analysis. A perceptual acoustical comparison of seven measures. JSHR, 29(1), 50-64.

Baynes, R.A. (1966). An incidence study of chronic hoarsenesses among children. JSHD, 31(2), 172-176.

Binu, R. (1998). An objective classification of voice disorders using artificial neural networks. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Biswajit. (1995). Multidimensional analysis of voice disorders in children. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Boone. D.E. (1983). "The voice and voice therapy", (2nd edition). New Jercy : Prentice-Hall Inc.

Boone. D.R. (1971). The voice and voice therapy. Second Edition, New Jercy : Printice Hall, Inc.

Boone. D.R. and Plante, E. (1993). Human communication and its disorders. New Jercy : Printice Hall, Inc.

Carhart. R. (1941). The spectra of model larynx tones. Speech monographs, 8, 76-84. Cited in Emanuel, F.W. (1991): Spectral noise. Seminars in Speech and Language, 12(2), 115-130.

Casper, M., Abramson, A.L. and Forman - Franco, B. (1981). Hoarseness in children: Summer camp study, Int. J. Pediatric Otorhinol., 3, 85-89. Cited in Sedereholm, E., McAllister, A., Sunberg, J. and Dalkvist, J. (1992). Perceptual analysis of child hoarseness using continuous scales. Speech transmission Laboratory Quarterly Progress and Status Report. 1, 99-133.

Coleman. R.F. (1960). Some acoustic correlates of hoarseness-unpublished master's thesis. Vanderbilt University, Nashville, T.N. Cited in Childrens, D.G. and Lee, C.K. (1991). Vocal quality factors: Analysis, Synthesis and perception. JASA. 90(5). 2394-2410.

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). Spectrographic Analysis of fundamental frequency and hoarseness before and after vocal rehabilitation. Journal of speech and Hearing Disorders, 39, 286-296.

David (1998). Acoustic correlates of hoarse voice. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Anitha, C. (1994). Multidimensional analysis of voice disosorders. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Fairbanks, G. (1960). Voice and articulation drillbook. New York. Harper and Bros.

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

Fluenj,Z., Minifie and Linn. (1992). Objective evaluation of pathological voices: a preliminary clinical clinician program. Paper presented at ASHA Convention, San Antoiro, November.

Gauffin, J., Hammarberg, B. and Hertegard, S. (1996). Irregularities in the voice: A perceptual experiment using synthetic voices with subharmonics. Cited in Davis, P.J. and Fletcher, N.H. (Eds): Vocal Fold Physiology. London : Singular publishing group.

Gould, W.J. (1975). "Quantitative assessment of voice function in microlaryngology". Folia Phoniatrica, vol. 27(3), 190-200.

Greene, M.C.L. (1964). The voice and its disorders. Whurr-AITBS, Delhi. Page 51.

Hartmann, E. and Cramon, D.V. (1984). Acoustic measurement of voice quality in central dysphonia. JCD, vol. 17(6), 425-440.

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

Heiberger, V.L. and Horii, Y. (1982). Jitter and Shimmer in sustained phonation. In Lass, N.J. (Ed) Speech and language: Advances in basic research and practice, vol. 7(pp. 299-332). New York : Academic Press.

Hirano, M. (1975). Phonosurgery: Basic and clinical investigations. Otologia (Fukuoka), 21, 239-240. Cited in baken, R.J. (1987), Clinical measurement of speech and voice. London: Taylor and Francies Ltd.

Hirano, M. (1981). Clinical examination of voice - Disorders of human communication, 5, Springer, Wien.

Hiraoka. N., Kitazoe, Y. Ueta, H., Tanaka, S. and Tanabe, M. (1984). Harmonic intensity analysis of normal and hoarse voices JASA, vol. 76(6), 1648-1651.

Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. JSHR, vol. 22(1), 5-19.

Imaizumi. S. (1986). Acoustic measures of roughness in pathological voice. Journal of phonetics, vol. 14,457-462.

Imaizumi. S., Hiki, S., Hirano. M. and Matsushita, H. (1980). Analysis of pathological voices with a sound spectrograph (in Japanese). J.Acoust. Soc. Jpn. 36. 9 - 16, Cited in Kasuya, H., Ogawa, S., Mashima, K. and Ebihara. S. (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice, JASA, vol. 80(5). 1329-1334.

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 Laryngology, vol. 76, 998-1007.

Jacob, I. (1968). A normative study of laryngeal jitter. Unpublished Master's Thesis. Univ. of Kansas.

Jayaram. K. (1975). An attempt at an objective method for differential diagnosis of dysphonics. An unpublished masters dissertation submitted to Mysore University, Mysore.

Kasuya. H., Masubuchi, K., Ebihara, S. and Yoshida. H. (1986). Preliminary experiments on voice screening. Journal of Phonetics, vol. 14. 463-468.

Kasuya. H. Ogawa, S. Mashima, K, and Soihara, S. (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. JASA. 80(5), 1329-1334.

Kasuya. H., Ogawa, S. and Kikuchi. Y. (1986). An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. Speech communication, 5-2.

Kasuya. M., Zue. W., and Endo, Y. (1993). Measurements of laryngeal turbulent noise in pathological voice. Paper presented at ASHA Convention. November, Aueherion.

Klatt, D.H., and Klatt L.C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. JASA, 87, 820-857.

Komiyama, S. and Buma. S. and Watanabe. H. (1973). Measurement of the phonation time taking account of the intensity of the voice. Otorhinolaryngology, Fukuoka, vol. 19(3): 385-389.

Kreiman. J., Gerratt, B.R., Kempster, G.B., Erman, A. and Berke, G.S. (1993). Perceptual evaluation of voice quality. Review, tutorial and framework for future research. JSHR, vol. 36(1), 21-40.

Kreiman, J., Gerratt B.R., Precoda, K. and Berke, G.S. (1992). Individual differences in voice quality perception. JSHR, vol. 35(3), 512-520.

Krishnamurthy, B.N. (1986). "The measurement of mean air flow rate in normals". An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Krishnan, B.T. and Vareed, C. (1932). The vital capacity of 103 medical students in South India. Ind. J. Med. Res, vol. 19, 1165-1183.

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

Leinonen, L., Hihunen, J. Laakso. M.L., Rihkanen, H. and Pop.ius, H. (1997). Categorization of voice disorders with six perceptual dimensions. Folia phoniatrica, vol. 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, vol. 35, 287-295.

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

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

Lieberman, P. (1961). Perturbation in vocal pitch. JASA. 33, 597-603. Cited in Askenfelt. A.G. and Hammarberg, B. (1986). Speech waveform perturbation analysis. A perceptual acoustical comparison of seven measures. JSHR, vol. 29(1),'50-64.

Lippman, R.P. (1987). Cited in speech information processing using artificial neural networks. In S.Nakagawa, K. Shikano, and Y. Tohkura (E&B). Speech Hearing and Neural Network Models (1995). Austerdam: Ohmsha IOS Press.

Lively, M.A. and Emanuel, F.W. (1970). Spectral noise levels and roughness severity ratings for normal and sknultated rough vowels produced by adult females. JSHR, vol. 13(3), 503-517.

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

Ludlow, C, Coulter. D., and Cardano, C. (1979). Application of pitch perturbation measures to the assessment of hoarseness in Parkinson's disease. Journal of Acoustic Society of America, vol. 66, 565-566.

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, vol. 25, 43-53.

Michael, I., Kierchner, F., Shelton, R and Mollinger, I. (1968). Acoustic results of teflon procedures Uni.Kan. Unpulished manuscripts.

Michael. J.F. and Wendahl. R. (1971). Correlates of voice production. In L.E. Travis (Ed.), Hand Book of Speech Pathology and Audiology (pp 465 - 480). New Jersey : Prentice-Hall. Inc. Englewood Cliffs.

Moore. G.P. (1971). Voice disorders organically based (Chapter-21). In L.E.Travis (Ed.), Handbook of speech pathology and audiology. New Jersey: Prentice Hall, pp.535-569.

Moore, G.P. (1971). Organic voice disorders. New Jersey : Prentice-Hall, Inc., Englewood Cliffs.

Moore, G.P. (1975). Observation on the physiology of hoarseness - Proceedings of the fourth international congress of phonetic science. Helsinki: Finland (pp. 92-95). Cited in Childrens, D.G. and Lee, C.K. (1991). Vocal Quality factors: Analysis, synthesis and perception. JASA, vol. 90(5), 2394-2410.

Moore, G.P. and Thompson, I. (1965). Comments on the physiology of hoarseness. Archieves of otolaryngology, 81. 97 - 102.

Mujunen, R., Leinonen, L., Kangas, J. and Tookkola, K. (1993). Acoustic pattern recogntiion of |sj misarticulation by self-organizing map. Folia phoniatrica, vol. 45 : 135-144.

Nataraja. N.P. (1986). Differential Diagnosis of Dysphonias. Doctoral Thesis, AIISH, University of Mysore.

Nataraja, N.P. and Jagadish. A. (1984). Vowel Duration and Fundamental Frequency. Journal of All India Institute of Speech and Hearing, vol. 15.

Patacek. P.H. and Sader, E.K. (1963). Maximum duration of phonation. Journal of Speech and Hearing Disorders, vol. 28, 171-182.

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

Pathak, C.B. (1995). The quantitative evaluation of hoarseness using harmonics to noise ratio and voice perturbations. An Unpublished Master's Dissertation submitted to Mysore University, Mysore.

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

Prabha. S. (1997). Analysis of dysphonic voice before and after therapy. Unpublished master's dissertation. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Preethi. (1998). Factors in normal and abnormal voice. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Qi, Y. and Shipp, T. (1992). An adaptive method for tracking voicing irregularities. JASA, vol. 91(6), 3471-77.

Rashmi, M. (1985). Acoustic aspects of the speech of children. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Sawashima, M. (1966). Measurement of Maximum phonation time. Journal of Logopedics and Phoniatrics, vol. 7, 23-28.

Scripture, E.W. (1906). Researches in experimental phonetics. The study of speech curvers. Washington, D.C. Carngegie Institute, (p.41). Cited in Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. JSHR, vol. 22(1), 5-19.

Sederholm, E., McAllister, A., Sundberg, J. and Dalkvist, J. (1992). Perceptual analysis of child hoarseness using continuous scales. Speech Transmission Laboratory Quarterly Progress and Status Report, vol. 1, 99-113.

Seth and Gruthrie (1935). Speech in Childhood.,London: Oxford University Press.

Shantha, Y.S. (1973). Establishing and Validating, Isochronal tone stimulation technique. An Unpublished Master's Dissertation submitted to the University of Mysore. Mysore.

Sheela, E.V. (1974). A comparative study of vocal parameters of trained and untrained singers. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

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

Simon, C. (1927). The variability of consecutive wave lengths in vocal and instrumental sounds. Psychological Monographs, 36, 41-83. Cited in Titze. I.R., Horii, Y. and Scherer, R.C. (1987). Some technical considerations in voice perturbation measurements. JSHR, vol. 30(2), 252-260.

Sonninen, A., and Sonninen, E. (1976). Logopedic-phoniatric observations on eight year old school children in the town of Jyavaskyla, Nord, Tidskr, Logopediog Foniatri.. 2, 74-82. Cited in Sederholm, E. McAllister, A., Sundberg, J. and Dalkvist, J. (1992). Perceptual analysis of child hoarseness using continuous scales. Speech Transmission Laboratory Quarterly Progress and Status Report vol. 1, 99-113.

Sreedevi, H.S. (1987). Acoustic characteristics of optimum voice. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

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. An Unpublished Master"s Dissertation submitted to the University of Mysore, Mysore.

Thaung, Z., Manifie, Ksasuya. H., and Lin. X. (1995). Measures of vocal function during change in vocal effort level. Journal of voice, vol. 9(4), 429 - 438.

Van Riper. C. and Irwin. J.V. (1978). Voice and Articulation. New Jersey : Prentice-Hall. Inc.

Vanaja, C.S. (1986). Acoustic Parameters of normal voice. An Unpublished Master's Dissertation submitted to the University of Mysore, Mysore.

Wolfe, V.I. and Steinfatt T.M. (1987). Prediction of vocal severity within and across voice types. JSHR, vol. 30, 230-240.

Yanagihara. N. (1967). Significance of harmonic changes and noise components in hoarseness, JSHR. vol. 10(3), 531-541.

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

Yanagihara. N. and Von Leden. H. (1967). Respiration and Phonation the functional examination of laryngeal disease. Folia phoniatrica, vol. 19, 153-166.

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

Yumoto, E. (1983). The quantitative evaluation of hoarseness: A new hormonics - to noise ratio method. Arch otolaryngol, vol. 109, 48-52.

Yumoto. E.. Gould. W.J. and Baer, T.(1982). Harmonics-to-noise ratio as an index of the degree of hoarseness, JASA. vol. 71(6), 1544-1550.

Yumoto, E.. Sasaki, Y. and Okamura. H. (1984). Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness.JSHR, vol 27, 2-6.

Zurada, J.M. (1992). Introduction to artificial neural systems. Boston: PWS Publishing Company.