Objective acoustic analysis of voice improvement after phonosurgery

Piyush Verma · Manisha Pal · Anoopraj

Abstract

Objective To evaluate voice improvement after phonosurgery by objective acoustic analysis.

Material and methods This prospective study was conducted in Maulana Azad Medical college New Delhi from December 2002 to 2008. In this study 100 subjects were included comprising of all age groups and either sex. All these patients with organic voice disorder underwent speech analysis using VAUGMI speech analysis program before and after phonosurgery.

Result All the parameter of voice analysis were deranged before the treatment but after surgery all the parameters should improvement most significant change was seen in the jitter.

Conclusion After doing this study we found that the analysis of hoarse voice using various parameters of acoustic analysis like fundamental frequency, harmonics to noise ratio, jitter, shimmer, S/z ratio helped us in identifying the degree of hoarseness and the severity related to it. Some parameters like jitter and shimmer were able to detect the component of hoarseness in perceptually normal voice and thus patient was helped by doing surgery and voice therapy at the appropriate time.

Keywords Phonosurgery · Voice analysis · Organic voice disorder

P. Verma¹ · M. Pal² · Anoop Raj¹

¹Department of ENT,

²Department of Paediatrics,
Maulana Azad Medical College
New Delhi, India

P. Verma (✉)

E-mail: doctor_piyushverma@yahoo.com

Introduction

Hoarseness is the cardinal symptom of laryngeal disease; considerable emphasis had been placed on this symptom. Several approaches, including acoustic, cinematographic aerodynamic and electrophysiologic had been utilized to explore the mechanism and pathophysiology of hoarse voice production. However, the evaluation of hoarseness (the estimation of the degree and the quality of hoarseness) was made chiefly on the basis of clinicians subjective perception. The basic protocol [1, 2] for functional assessment of voice pathology, especially for assessing the efficacy of treatment included:

Perception, which included roughness and breathiness

The severity of hoarseness was quantified under the parameter G (Grade) for the GRABAS scale [3]. The main components of hoarseness were:

- Roughness or hoarseness (R): Audible impression of irregular glottal pulses, abnormal functions I fundamental frequency and separately perceived acoustic impulses due to irregular vocal fold vibrations. The other parameters like asthenicity (A) and stream (S) were less reliable and were omitted.
- Breathiness (B): Audible impression of turbulent air leakage through an insufficient glottic closure, which included short aphonic movements.

Grading scale was made for the reporting purposes:

- Normal or absence of deviation
- Slight deviation
- Moderate deviation
- Severe deviation.

Video stroboscopic evaluation

It had been the main clinical tool for etiological diagnosis of voice disorders. It was also used assessing the quality

of vocal fold vibrations and thus studying the effectiveness of the treatment. Basic parameters of stroboscopy included-glottal closure.

Longitudinal: Over the whole length of the glottis and without sufficient adduction.

Dorsal: Posterior and triangular chink (it normal in 60% of the middle aged healthy women) ventral irregular.

Oval: Over the whole length of the glottis but with a dorsal closure hourglass shaped [11].

Rating of the glottal closure was found to be very reliable regularity:

Quantitative rating of the degree of irregular slow motion. Mucosal wave/quantitative rating of the quality of the mucosal wave, accounting for the physiology of the layered structures of the vocal folds [12].

Symmetry: Quantitative rating of the mirror motion of both the vocal cords.

Aerodynamics, which included: Phonation quotient or the maximum phonation time.

Acoustic measurement [4, 5], which included: Fundamental Frequency/Fo, Harmonics to Noise Ratio, Jitter, Shimmer, etc.

Subjective evaluation

Acoustic analysis: It provided objective and non-invasive measures of vocal function.

Fundamental frequency/Fo [6] – It is the number of vibrations of the vocal folds per second. Fo is the habitual pitch most often used by the individual. The normal value for males lies between 80–180 Hz and for females it is 180–280 Hz. A high-pitched voice is seen in spastic dysphonia, puberphonia and in laryngeal webs. A low-pitched voice is seen in case of vocal fold thickening or mass on vocal folds.

Fundamental frequency range – It is the difference between the lowest and the highest frequency. The normal range lies between 1–1.5 octaves for males and 2–2.5 octaves for females.

Optimal frequency – It is the best suitable frequency for an individual which is marked maximum loudness with minimum physiological effort. It can be labeled as a frequency which is 1/3 of the frequency range. Ideally Fo = Optimal frequency.

Harmonics Harmonics-to-noise (H/N) ratio [7, 8, 16] – when the vocal folds are in the state of vibration, besides generating fundamental frequency, they also generate harmonics which are multiples of fundamental frequency. The energy of harmonics decreases as their frequency increases. Noise is produced when there is no firm closing of the glottis due to which there is turbulent air escape. The ratio of harmonics to noise intensity should always be >1.

In conditions of slackness of vocal folds seen in paralysis, nodule or reduced glottic chink, the noise is greater causing a reduction in H/N ratio.

Number of harmonics – This parameter can be determined by long term average spectrum (LTAS) at low frequency (0–2000 Hz). Number of harmonics visible should be >18 on LTAS. In case of increased noise level, the noise masks that higher harmonics resulting in a decrease in the number of visible harmonics.

A study was conducted by Nieto et al. [9] in 1996 in which LTAS and HNR were taken as the basic parameters for preoperative and postoperative evaluation of voice. Before surgery LTAS revealed weak harmonics. After surgery H/N values and harmonic energy was shown to be increased. Preoperative and postoperative changes were significant ($p < 0.05$). Improved glottic closure after surgery improved both the LTAS and HNR parameters.

S/z ratio – Gamboa et al. [10], in 1995 conducted acoustic analysis by taking S/z ratio as the basic parameter for 72 patients with laryngeal pathology. This is defined as the ratio of the maximum phonation time for which sound/s/ can be sustained for the maximum phonation time for which that sound/z/can be phonated. This study concluded that in a normal speaking subject it was nearly 1.0. When glottic closure defect was present, the ratio was over 1.4. An organic lesion where there was difficult glottic closure, there was decreased airflow resistance and therefore, a shortened phonation time of vowel and thus increased S/z ratio.

Jitter – It is the cycle to cycle variation in frequency during vibrations of the vocal folds. Variation in frequency by 3 Hz is considered to be normal.

Shimmer – It is the cycle to cycle variation in intensity. Variation in intensity over 3 dB is considered normal.

Small variations (perturbations) in amplitude and period-time from cycle to cycle in the speech waveform are known to be the natural ingredients in normal speech (Liberman, 1961). In fact, such perturbations were important for the natural quality of speech synthesis (Holmes, 1962). These variations in pitch and amplitude were probably due to the periodicity of the neuromuscular phonatory control system (Shultz-Coulon, Batamar and Fedder, 1979). It was also known that large magnitude perturbations gave rise to a “rough” voice quality (Coleman, 1971). Large perturbations reflected alterations in the normal pattern of vocal cord vibrations (von Leden, 1960), and were associated with laryngeal dysfunction (Hecker, 1971).

Liberman in 1961 performed pitch perturbation analysis and defined a pitch perturbation factor as the percentage occurrence of pitch period perturbation larger than 0.5 m in a vocal segment of a connected speech and this perturbation factor was useful in determining the laryngeal diseases.

Koike in 1973 studied voice of 30 subjects with laryngeal diseases and 30 normal subjects. In contrast to Liberman's

measurements, in which perturbations exceeding the fixed absolute factor were included, Koike registered the relative pitch perturbation. Using the measures on sustained vowels, he was able to discriminate between patients with tumors, vocal cord paralysis, and normal subjects.

Rabinov et al. [11] conducted a study, in 1995 comparing the reliability of perceptual rating of roughness with acoustic measure of jitter. This parameter was taken for comparison because it was extensively studied, widely available in commercial kits and easy to compare and calculate. The results suggested that acoustic measures of jitter had advantages over perceptual measures of roughness for discriminating among normal voices, that is, measures of jitter may resolve very small aperiodicities that are presumably near or below listener's perceptual thresholds. This study also suggested that measured jitter was the function of both signals and algorithms.

Subjective evaluation: This evaluation of voice although subjective was very important. The basic aim was to differentiate the deviance of voice quality and the severity of the disability in daily and professional life. A voice handicap index [12] was computed on the basis of patient's response to a carefully selected list of questions. A score of '0' (extreme left) meant normal voice (no deviance) on the first scale and no disability or handicap (related to voice) in daily life on the second scale, while '100' (extreme right) meant extreme voice deviation on the first scale and extreme disability in daily social activities, as rated by the patient himself.

A new handicap index called the voice handicap index-10 [13] which is an abbreviated form of voice handicap index has been developed.

Voice related quality-of-life (V-RQOL) 14 scale comprises 10 items divided into physical functioning and social-emotional functioning subscale. Each item is scored on a five point interval scale that reflects the degree of problem. For the subscales and the total score, 100 is the highest possible score and reflect the highest quality-of-life.

Another index called the glottal function index [17] is being used these days which is a simple self administered 4-item battery, which is specifically aimed at identifying the presence and the degree of symptoms of glottal dysfunction. The scale ranges from 0 (asymptomatic) to 20. The questions included like speaking taking extra effort, throat discomfort while speaking, vocal fatigue, voice cracks.

Movement of vocal folds and anatomy of speech

The vocal folds [15] usually vibrate at 100–300 Hz during normal conversation and even at 1000 Hz or even more during singing. In quiet respiration, the inter-membranous part of the glottis is triangular and the inter-cartilaginous

part is rectangular as the medial surfaces of the arytenoids are parallel.

In forced respiration, the vocal folds undergo extreme abduction; the arytenoids cartilages are pulled laterally and their vocal processes are pulled widely apart, changes in length and tension of the vocal folds control the pitch of the voice and are produced normally only when the vocal folds are in contact for phonation. Three following forces act to bring the vocal folds in contact with each other.

- I Tension in the folds
- II The decrease in sub-glottic air pressure
- III The sucking effect of the escaping air (Bernoulli's effect).

The result of the rapidly repeating cycles of opening and closing at the glottis releases small puffs from the sub-glottic air column, which form sound waves. Frontal tomography showed that the areas of vocal fold surface in contact vary in according to pitch. At low pitch the cross-sectional area of the vocal fold increases but as the pitch rise the vocal folds become thinner.

Stroboscopy allows observation and description of fundamental frequency, symmetry of bilateral movement of the vocal fold, regularity, glottal closure, amplitude, mucosal wave and non-vibrating portion. In the normal vocal folds the mucosal wave travels on the mucosa from its inferior to superior surface. This is observed during vibrations except for falsetto voice and is a function of soft and pliant superficial layer of the lamina propria. The function of the vocal folds is to produce sounds varying only in intensity and pitch. This is modified by various resonating chambers above and below the larynx and is ultimately converted to phonemes by the articulating action of the pharynx, tongue, palate, teeth and lips.

Spectral analysis shows that the vocal tract (larynx. Pharynx, mouth and nasal cavity) acts as an intricately selective filter and resonator, which propagates a remarkably similar pattern irrespective of fundamental frequency. This is essential to speech as it ensures that in spite of a continuous variation in tone, a constant quality of timbre is maintained.

Materials and methods

The objective acoustic analysis of voice improvement after phonosurgery was conducted in the Department of Otorhinolaryngology and Head and Neck surgery, Maulana Azad Medical College and Lok Nayak and associated hospitals from December 2002–2008. A total of 100 patients were selected for various phonosurgical procedures like micro laryngeal surgery and medialization thyroplasty

and the results were compared between preoperative and postoperative voice evaluation.

Patients

Patients with organic voice disorders like vocal nodules, vocal polyps, Rinkes edema, and unilateral vocal cord paralysis were included in the study. Patients with functional causes of voice change and malignancy of the larynx were excluded from the study.

Materials

Operating equipment for microlaryngeal surgery and thyroplasty speech analysis software program - VAGHMI, digital tape recorder personal computer with windows 98, 2000, Me, Xp, and other hardware like amplifier, microphone, and laser printer.

Methods

In this prospective study all patients were taken up for objective acoustic analysis, before and after phonosurgery, by VAGHMI software program. Audio recording was the most useful basic tool for voice assessment. The high quality voice recordings were made on digital tape recorder and were later subjected to acoustic analysis. All the recordings were done in a quiet room with ambient noise of <50 db. Voice sampling was done at a frequency of 2000 Hz. The mouth to microphone distance was set constant at 10 cm. Patient was asked to take a deep breath and say /a/ without straining.

All the microlaryngeal surgery were done under general anesthesia and type I thyroplasty were done under local anesthesia. Patients were given antibiotics and decongestants and were advised voice rest for 1 week postoperatively. Speech therapy was started after 1 week postoperative and patient were called every fortnightly for follow up. Speech analysis using VAUGHMI software was done all the patients after 1 month.

Observations and results

- **Patient profile:** Hundred patients were included in this study. The age of the patient varied from 10–55 years with a mean age of 33 years.
Number of males - 62.
Number of females - 38.
- **Etiology:** This study included cases like benign vocal cord lesions like polyps, nodule, papillomas or

unilateral vocal cord palsies.

Vocal	polyps	38
Vocal	s nodule	22
Juvenile laryngeal papillomatosis		12
Idiopathic vocal cord palsy		08
Rinke's edema		02
Vocal cord thickening		18

- **Voice analysis:** It was done by a commercially available software program by the name of VAUGHMI. Statistics were done using SPSS-10 software and using student t-test.
- **Fundamental frequency/Fo (in Hz):** In this study it was observed that preoperative frequency was higher than normal or on the higher side of normal. After surgery and speech therapy this parameter was found to be in the normal range.

Preoperative	Mean: 207	Std. Dev.: 42.35
Postoperative	Mean: 144.1	Std. Dev.: 23.86
Correlation coefficient	0.439	p value 0.15

- **Frequency range:** It is the difference between the lowest and the highest frequency one uses while phonating or speaking. While reading the range should be from 25 to 30 Hz. This parameter was abnormally high pre-operatively but has improved after surgery.

Preoperatively	Mean: 56.8	Std. Dev.: 84.08
Postoperative	Mean: 11	Std. Dev.: 4.38
Correlation	0.358	p value 0.52

- **Optimal frequency:** It is the best suitable frequency an individual that is marked by a maximum loudness with minimum physiological effort. It can be labeled as a frequency, which is 1/3 of the fundamental frequency. In this study there was an improvement in this parameter after surgery as there was improvement in fundamental frequency.

Pre-operative	Mean: 202.7	Std. Dev.: 38.30
Post-operative	Mean: 143.2	Std. Dev.: 19.50
Correlation	0.465	p value 0.014

Harmonics to noise ratio, H/N: When vocal cords are in a state of vibration they generate harmonics which are multiples of fundamental frequency. In hoarse voice they are replaced by irregular pattern of noise. The normal value for H/N 16 ratio for males and females is >13. In this study it was observed that H/N ratio was low which meant that noise component was more than the harmonic component, but postoperatively it increased showing that harmonic component had improved, but it was not possible to find the diagnostic correlation between H/N ratio and the type of pathology causing hoarseness; however it was seen that H/N ratio was related to the structural changes to the vocal cords and thus it was noticed that the ratio was low in vocal

cord palsy and in cases of Rinkie's edema.

Preoperative	Mean: 11.8	Std. Dev.: 2.4
Postoperative	Mean: 17.5	Std. Dev.: 2.15
Correlation	0.266	p value 0.155

4. **Jitter (%):** It is the cycle-to-cycle variation in frequency during vibration of the vocal fold. The normal range is 0.33 to 7.52. It was above the normal values before surgery and became normal after surgery.

Preoperative	Mean: 14.3	Std. Dev.: 9.49
Postoperative	Mean: 6.5	Std. Dev.: 2.59
Correlation	0.586	p value 0.001

5. **Shimmer (dB):** It is the cycle-to-cycle variation in intensity. The normal values ranges from 0.12 to 1.12 dB. In our study it was found to abnormal or higher side of normal. But after surgery it was found to be within the normal range.

Preoperative	Mean: 0.76	Std. Dev.: 0.34
Postoperative	Mean: 0.49	Std. Dev.: 0.15
Correlation	0.371	p value 0.043

Number of harmonics: It is the number of visible peaks in the LTAS graph. It should be >18 to be considered to be within normal limit. In case of increased noise level, the noise masks the higher harmonics resulting in a decrease in the number of visible harmonics. In this postoperative period after receiving voice therapy the number of harmonics has increased.

Preoperative	Mean: 5.6	Std. Dev.: 1.73
Postoperative	Mean: 13.1	Std. Dev.: 2.21
Correlation	0.194	p value 0.30

6. **S/z ratio:** It is the ratio of time for which the sound /s/ can be sustained to the sound/z/can be phonated. Normal value ranges from 0.9 to 1.1sec. In this study this parameter improved with surgery in most of the patients.

Preoperative	Mean: 1.27	Std. Dev.: 0.15
Postoperative	Mean: 1.04	Std. Dev.: 0.12
Correlation	0.149	p value 0.43

7. **Maximum phonation duration:** It is the time for which one can continue phonation after a deep breath. The normal range is from 12 to 15 secs. It was normal in both preoperative and postoperative periods, but it improved to higher values after surgery. The values

were really for patients with unilateral vocal cord paralysis.

Preoperative	Mean: 12	Std. Dev.: 2.16
Postoperative	Mean: 15.3	Std. Dev.: 2.12
Correlation	0.112	p value 0.55

Discussion

In this study of 100 patients, with age ranging from 10–55 years with a mean age of 33 years, various organic causes of hoarseness like vocal nodule and polyp, Rinke's edema, vocal cord palsy and papillomatosis were taken into consideration. There were total of 38 female patients and 62 male patients including 1 to 10-year-old girl and 1 to 10 year old boy. The reason for this discrepancy regarding sex distribution could be attributed to the fact that males are more involved in outdoor activities and smoking and drinking is also more prevalent in this sex.

All the 100 patients underwent objective acoustic analysis both preoperatively and postoperatively. Various acoustic parameters measured were fundamental frequency, frequency range, jitter, shimmer, etc. 92 patients underwent microlaryngeal surgery for various benign vocal fold lesions and 08 patients underwent medialisation thyroplasty (Ishikiki type I) who were diagnosed with idiopathic unilateral vocal cord paralysis. After 4 weeks of surgery, during which patient was given voice therapy, acoustic analysis was repeated again for the same parameters.

It was observed that before surgery and voice therapy all the parameters of voice like fundamental frequency, optimal frequency, jitter, S/z ratio, shimmer, harmonic to noise ratio, number of harmonics, were having abnormal values; but after treatment and voice therapy all the values were found to be within normal values but only jitter was statistically significant.

In our study it was observed that fundamental frequency was relatively high for patients with benign vocal cord lesions than for vocal cord palsy. This finding could perhaps be attributed to the fact that, due to the non-closure of the glottic chink there was escape of air and thus decreased vocal cord vibration. Optimal frequency was also decreased after the surgery. Various cases like laryngeal papilloma had comparatively less postoperative changes as compared to vocal nodule or medialization thyroplasty similarly was the case with Rinke's edema (Fig. 1).

The results showed that there were abnormally high values of jitter and shimmer before surgery, this could be attributed to the fact that, when there was a sustained phonation, there was relatively longer open phase during the vibratory cycle of the vocal cords and a shorter closed phase and thus increased vocal fold vibration leading to

Fig. 1 Comparison of preoperative and postoperative acoustic analysis

these abnormal values. These changes in perturbation values could also be due to the contraction of the vocalis muscle, which is the primary agent for monitoring the stable phonation. It might be hypothesized that as the frequency and intensity increases, the muscular forces exerted by the vocalis and cricothyroid muscle gets misbalanced and thus causing unstablising effect leading on to the change in the vibratory pattern of the vocal folds.

Acoustic analysis is a convenient, non-invasive method of detecting aperiodicity in the speech wave form, but it does not give definitive information concerning the physiologic causes of aperiodic performance of the vocal organ. The nature of aperiodicity must be determined primarily by direct observation of the larynx. There is a close connection between the degree of perceived roughness and degree of irregularity in the vocal fold vibration, that is, the degree of irregularity increased with the increased amount of perceived roughness in voice. Because direct examination of the larynx could interfere with the normal phonation and causes discomfort to the patient, it is always better to utilize acoustic measures for hoarseness.

Summary and conclusion

The aim of the study was objective acoustic analysis of voice after phonosurgery. The analysis was done before the surgery and after 4 weeks of surgery during which time speech therapy was given to the patient. After doing this study we found that the analysis of hoarse voice using various

parameters of acoustic analysis like fundamental frequency, harmonics to noise ratio, jitter, shimmer, S/z ratio helped us in identifying the degree of hoarseness and the severity related to it. Some parameters like jitter and shimmer were able to detect the component of hoarseness in perceptually normal voice and thus patient was helped by doing surgery and voice therapy at the appropriate time.

As we found the advantages of this acoustic analysis, we found the shortcomings of this measurement; that it could not predict the pathology causing dysphonia which could only be confirmed on direct visualization of the vocal cords by either doing direct or indirect laryngoscopy; but direct examination of the larynx could interfere with the normal phonation and causes discomfort to the patient, thus it is always better to utilize acoustic measures for hoarseness along with the other investigations.

Voice analysis should be done on regular basis as part of the routine workup of the patient with hoarse voice, but it cannot substitute other parameters like perceptual analysis of voice, videostroboscopy, and direct and indirect laryngoscopy. It can work as complementary test for voice and mainly for research purposes, the main outcome will always be subjective evaluation voice by the patient himself.

References

1. Fresnel E, Fourcin AJ (2000) Rev Laryngol Otol Rhino (Bord). 121(5):315–318
2. Phillippe HD, Patric B, Pats C (2000) A basic protocol for

- functional assessment of voice pathology, especially for investigating the efficacy of phonosurgical treatments and evaluating new assessment techniques: *Eur Arch Otorhinol* 258:77–82
3. Dejonckere PH, Remacle M, Fresnell EE (1996) Differentiated perceptual evaluation of pathological voice quality; reliability and correlation with acoustic measurements. *Rev Otorhinol* 117:219–224
4. Hirano M, Jibi S, Yoshida T (1998) Acoustic analysis of pathological voice: *Acta Otolaryngol (Stock)* 105:432–438
5. Nieto A, Del Palacio AJ, Lorezo FJ (1995) Computer use in voice analysis: Clinical applications: *Acta Otorhinol Esp* 46(4):214–245
6. Maryloy P (1995) Fundamental frequency, intensity and vowel selection; effect on measures of phonatory stability. *J Speech Hearing* 38:1189–1198
7. Nieto A, Cobeta I, Gamboa FJ (1996) Harmonics to noise ratio and spectrographic analysis in vocal abuse pathology. *Acta Otorhinol Esp* 47(5):370–376
8. Giovanni A, Revis J, Triglia JM (1999) Voice improvement after phonosurgery. *Laryngoscope* 109
9. Nieto, et al. (1996) Harmonics to noise ratio and spectrographic analysis in vocal abuse pathology. *Acta Otorhinol Esp* 47(5):370–376
10. Gamboa FJ, Nieto A, del Palacio (1995) S/z ratio in glottic closure defects. *Acta Otorhinol Esp* 46(1):45–48
11. Rabinov, et al. (1995) Comparing reliability of perceptual rating of roughness and acoustic measures of jitter and shimmer. *J Speech Hearing Res* 38:26–32
12. Hirano M, Jibi S, Yoshida T (1998) Acoustic analysis of pathological voice: *Acta Otolaryngol (Stock)* 105:432–438
13. Rosen CA, Lee AS, Osborne J, Zullo T (2004) Development and validation of the voice handicap index-10. *Laryngoscope* 114(9):1549–1556
14. Hogikyan ND, Sethuraman G (1999) Validation of an instrument to measure voice related quality of life (V-RQOL). *J Voice* 13:557
15. Hammarberg B, Fritz G, Sunderberg J (1980) Perceptual and acoustic correlation of abnormal voice quality: *Acta Otorhinol* 90:441–451
16. Yumoto E, Sasaki Y, Okamoto H (1989) Harmonics to noise ratio to measure the degree of hoarseness. *J Speech Hearing Res* 27:2–10
17. Kevin KB, Peter CB, Kathleen W (2005) Validity and reliability of the glottal function index: *Arch Otolaryngol Head Neck Surg* 131:961–964