

Kirt Areis E. Delovino, MD  
Ray U. Casile, MD  
Frederick Y. Hawson, MD

Department of Otorhinolaryngology  
Head and Neck Surgery  
St. Luke's Medical Center

## Vocal Acoustic Measures of Asymptomatic Filipino Young Adults at a Private Tertiary Hospital in Quezon City – A Pilot Study

### ABSTRACT

**Objectives:** To describe the vocal acoustic measures of non-smoking Filipino young adults without voice complaints at a private tertiary hospital in Quezon City; to determine if our baseline values are distributed normally and comparable to data in similar studies done abroad; and to recommend normative voice parameters which may be used as baseline data in our institution and for comparison in future studies.

### Methods:

**Design:** Cross-sectional study

**Setting:** Private tertiary hospital

**Participants:** A total of 70 subjects were recruited at random

**Results:** Values extracted for  $f_0$ , Jitter %, Jitter dB, Shimmer %, Shimmer dB and NHR showed normal distribution of results. The average vocal acoustic values found in the present study for male voices producing the vowel /a/ were  $f_0 = 130.6 \pm 13.65\text{Hz}$ , jitter =  $0.046\% \pm 0.184$ , jitter dB:  $37.62\text{dB} \pm 16.664$ , shimmer % =  $0.23\%$ , shimmer dB =  $0.23 \pm 0.67$  and NHR =  $0.13 \pm 0.010$ . The average values found for female voices, producing the vowel /a/ were  $f_0 = 218.38 \pm 26.192\text{Hz}$ , jitter =  $0.87\% \pm 0.61$ , jitter dB:  $34.82 \pm 22.5$ , shimmer % =  $2.72 \pm 1.07$  shimmer dB =  $0.23\text{dB} \pm 0.67$  and NHR =  $0.12\text{dB} \pm 0.016$ . Values retrieved from this study show similar trends with other papers abroad.

**Conclusion:** Voice acoustic systems are composed of different recording criteria, recording instrumentations and algorithms which primarily cause the differences in the results obtained in various studies, thus, precluding a single normalization. Following international recommendations for individual normalization per institution, we have obtained our own values. Our data was comparable to the results of other international studies. However, further investigation is recommended in areas where possibilities of interdialectic variation may produce an effect on the outcome of the study.

**Keywords:** *vocal acoustic measures, computerized speech lab, normative voice parameters*

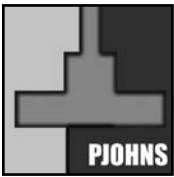
**In the Philippine setting,** voice and speech problems are often initially assessed by Ear, Nose and Throat (ENT) specialists. These evaluations are generally gauged subjectively by means of hearing perception. Perceptual evaluation of voice is an important scientific process in clinical investigation and in assessing voice quality, relevant deficiencies and their effect on the subject's ability to communicate. However, perceptual evaluation has some restrictions because of its poor correlation between evaluators.<sup>1</sup> Moreover, there exists a number of scales and their reliability varies from study to study. These limitations lead to numerous differences and non-standardization.

Correspondence: Dr. Kirt Areis E. Delovino  
Department of Otorhinolaryngology  
Head and Neck Surgery  
St. Luke's Medical Center  
279 E. Rodriguez Ave, Quezon City 1102  
Philippines  
Fax: (632) 723 0101 local 5543  
E-mail: edot\_ii@yahoo.com  
Reprints will not be available from the author.

The authors declared that this represents original material that is not being considered for publication or has not been published or accepted for publication elsewhere in full or in part, in print or electronic media; that the manuscript has been read and approved by all the authors, that the requirements for authorship have been met by each author, and that each author believes that the manuscript represents honest work.

Disclosures: The authors signed disclosures that there are no financial or other (including personal) relationships, intellectual passion, political or religious beliefs, and institutional affiliations that might lead to a conflict of interest.

Presented at the Descriptive Research Contest, Philippine Society of Otolaryngology Head and Neck Surgery, Glaxo Smith Kline (GSK) Bldg., Chino Roces Ave., Makati City, Philippines, October 11, 2010.



Over the past decade, an increasing number of studies have aimed at different objective analyses of vocal acoustics. Among these tools, computer-based acoustic analysis has become a more popular system in studies intended for the objective assessment of vocal parameters. A lot of these researches were intended to establish parameters necessary to create normal and standard values.

Many acoustic parameters of the human voice are evaluated by these computer systems. The most common parameters used in voice assessment in the literature are: fundamental frequency (**f<sub>0</sub>**), cycle-cycle perturbations such as jitter (**jitt**) and shimmer (**shim**) and the noise-harmony ratio (**NHR**).<sup>2,3,4,5</sup>

The fundamental frequency is an important parameter in both functional and anatomical larynx assessment.<sup>4,6</sup> It is determined by the number of cycles produced by the vocal folds per second and is reflective of the interaction of vocal fold length, mass and tension during speech.<sup>3</sup> Among acoustic parameters, fundamental frequency has been proven to have higher uniformity among different acoustic analysis systems and is less sensitive to voice recording characteristics.<sup>3,4</sup>

During phonation wherein there is sustained vibration of the vocal folds, there are occasional slight variations of the vocal folds' regular oscillation from cycle to cycle, otherwise termed as perturbations. These phenomena are called frequency perturbation (jitter) and amplitude perturbation (shimmer). These two correlate with the subjects' degree of roughness.<sup>2,6</sup>

The noise-harmony ratio characterizes the relationship between the two components of the acoustic wave of a sustained vowel: 1) the periodic component, which are the vocal fold regular sign and the additional noise coming from the vocal folds; and 2) the vocal tract.<sup>3,7</sup>

At present, there are several different automated vocal acoustic analysis systems and each system provides consistent, reliable and repeatable results in extraction of fundamental voice parameters. However, uniformity between these systems varies considerably. Felipe *et al.*<sup>3</sup> recommended establishing and normalizing vocal parameter values individually as their values differ considerably. Thus it is necessary to normalize the data from the software we are utilizing.<sup>2,3,4</sup>

Instrumental measures of the vocal function form an integral component of the clinical process in institutions abroad, rather than a supplement to assessment and treatment.<sup>9</sup> Objective acoustic analysis will certainly add more accuracy and impartiality in the evaluation of dysphonic patients resulting in more scientific management. Aside from providing an objective measure, this noninvasive procedure would also present adjuvant approaches to dysphonia and allow reliable comparison of voice samples (e.g., before and after treatment), therapeutic methods (e.g., microsurgery versus laser), or surgical groups. These measurements may also serve to provide baseline data in monitoring the degree of improvement in patients undergoing voice training as well as those in speech rehabilitation.

The local paradigm of treatment in patients with voice and

speech problems usually involves initial otolaryngologic evaluation. Management is usually based on laryngoscopic examinations. Acoustic examination is seldom considered unless long term therapy is required. In most common subtle symptoms of dysphonia (observed in singers with difficulty reaching habitual pitches), this measure may offer assistance in the diagnosis and treatment.

The goal of the present study is to describe the vocal parameters fundamental frequency, jitter, shimmer and noise-harmony ratio (NHR) measures for the CSL 4400 software, from Kay Elemetrics, used in the Voice Analysis Laboratory of a private tertiary hospital in Quezon City and to determine if our data is comparable to international studies.

## METHODS

### Study Design

This was a quantitative cross-sectional descriptive study. This study was approved by the Ethics Committee for Research of the Department of Otorhinolaryngology, St. Luke's Medical Center, and informed consent was obtained from all participants.

### Setting

Data collection was carried out in a sound treated room at the Voice Analysis Laboratory of the St. Luke's Medical Center.

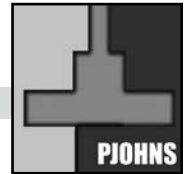
### Participants

A total of 70 subjects were recruited at random by one of the authors from among department consultants, resident physicians, nurses and other hospital employees, medical interns and clinical clerks. As a pilot study, the number of proponents was set arbitrarily in consonance with international literature.

Inclusion criteria were age between 20 and 45 years, absence of any signs and symptoms of voice change and no smoking history.<sup>7</sup> Exclusion criteria were: recent history of altered voice performance, voice complaints such as hoarseness, voice fatigue, voice failure or irritated throat since these symptoms suggest organic alterations of voice that might affect study results;<sup>14</sup> common cold, sore throat or upper respiratory tract infections since these conditions may cause phonation apparatus edema and dysfunction - or other diseases that could limit voice production during the evaluation; or any prior voice therapy or professional voice training and/or otorhinolaryngologic treatment as these subjects may consciously alter self-monitoring of voice and compromise voice quality. Singing in choirs or professional singers was also excluded to avoid subjects with trained voices.

### Data Gathering and Sampling Procedure

After giving informed consent, the subjects were given a data checklist to be answered completely to assess the selection criteria and afterwards interviewed for history-taking. Aside from not presenting voice alterations signs and symptoms (from data checklist and history-



taking), the participants' voices were also screened using the GRBAS system by a speech therapist who worked with the Speech Rehabilitation Clinic in the same hospital and an otolaryngologist/vocologist who was also an author of the study. Only data from the individuals considered with normal voice were included in the study.

Data collection was obtained using the Multi-Dimensional Voice Program software with Computerized Speech Lab CSL- Model 4400 from Kay Elemetrics (KayPENTAX, Montvale New Jersey, USA). Coupled with the CSL Kay Elemetrics model 4400 Digital recorder, a hi-fidelity microphone was used, Sennheiser model E 815 S (Sennheiser Electronic Corporation, Lyme, Connecticut, USA) and it was kept at a fixed distance of 5 cm in front of the subject's mouth.<sup>3,7,8,10</sup> The subjects were seated facing away from the monitor to prevent self-monitoring and conscious alteration of their voices during sampling. We used the sustained vowel /a/ at a habitual frequency and intensity following a deep breath, issuing the sound to achieve maximum phonation time without using expiratory reserve air. In order to stimulate habitual pitch and loudness, the subjects were also asked to utter a phrase immediately prior to the sustained vowel. The sustained vowel is preferred over regular speech in vocal acoustic assessment as it provides more reliable results.<sup>10</sup> A total of five samplings were done for at least 3 seconds each. The first two samples were excluded to avoid voice onset effects on data analysis. Vocal intensity was controlled by monitoring the software's Vu meter. When the sample exceeded the software's acceptable Vu range, a new sample was collected.

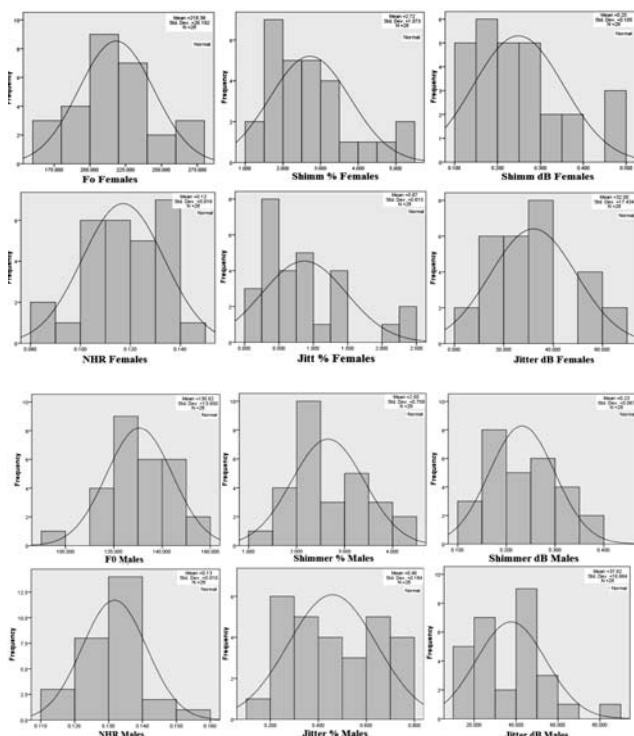


Figure 1. Kolmogorov-Smirnov Test - Histogram showing normal distribution curves per parameter among male and female subjects.

Tests of Normality for Female Values

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistic	df	Sig.
F0	.096	28	.200 <sup>*</sup>
Shimmdb	.135	28	.200 <sup>*</sup>
shim	.121	28	.200 <sup>*</sup>
jitterdb	.126	28	.200 <sup>*</sup>
jitter	.150	28	.110
NHR	.088	28	.200 <sup>*</sup>

a. Lilliefors Significance Correction

\*. This is a lower bound of the true significance.

Tests of Normality for male values

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistic	df	Sig.
F0	.101	28	.200 <sup>*</sup>
shimmdb	.142	28	.153
shim	.138	28	.183
jitterdb	.123	28	.200 <sup>*</sup>
jitter	.128	28	.200 <sup>*</sup>
NHR	.119	28	.200 <sup>*</sup>

a. Lilliefors Significance Correction

\*. This is a lower bound of the true significance.

Figure 2. Kolmogorov-Smirnov Test for Normality – computed adjusted scores with application of Lilliefors correction. All parameters for male and female subjects revealed normal distribution.

The voice samples were studied based on following acoustic parameters: fundamental frequency (Hz), jitter (%), absolute jitter (dB), shimmer (%), absolute shimmer (dB) and noise-harmony ratio (NHR). Each of these parameters was also analyzed as to gender. The descriptive statistical data analysis was carried out through SPSS for Microsoft Windows Version 16.0 (IBM Corporation, Armonk, New York, USA). Data were assessed statistically by applying descriptive statistics. The Kolmogorov-Smirnov method was applied for assessing the normality of results; the significance level was set at 5% ( $p > 0.05$ ); this yielded a results distribution curve (Figure 1) and was applied for Normality testing. (Figure 2)

RESULTS

A total of 56 young adults (28 men and 28 women) met inclusion criteria and participated in this study. Their ages ranged between 22-43 years old (mean 29) and included hospital employees, nurses, medical clerks and interns, consultants, resident physicians and staff

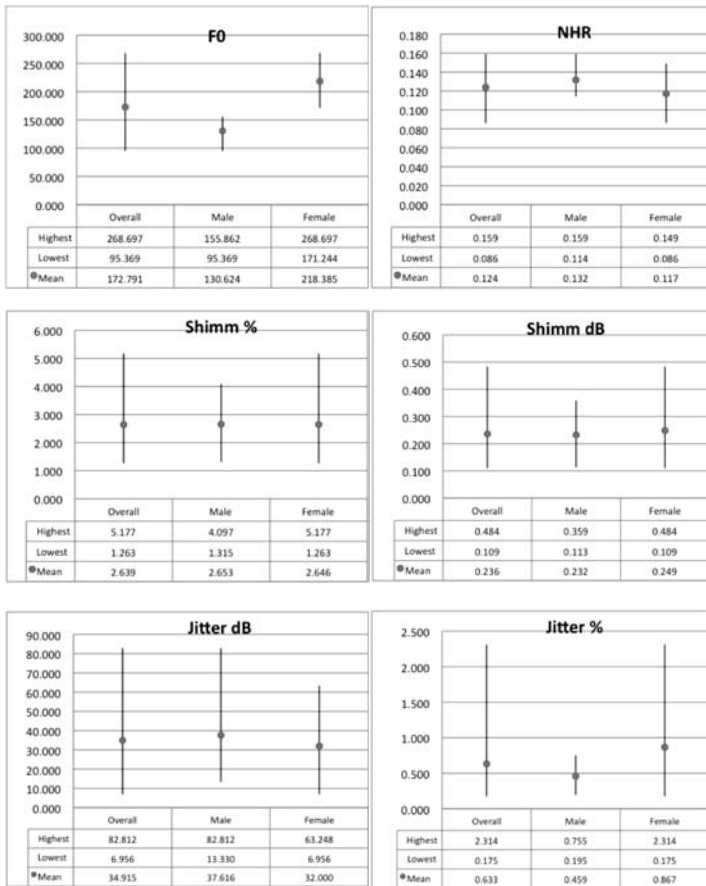


Figure 3. Test results: Stock chart graph showing individual parameter results overall, and in male and female groups.

who worked at the Voice Analysis Laboratory of the hospital. Of the 70 initially recruited for the study, four were excluded due to recent-onset upper respiratory tract infection with noticeable voice changes, six had just quit smoking for less than a month and another four refused to enroll in the study.

Values extracted for f0, Jitter %, Jitter dB, Shimmer %, Shimmer dB and NHR showed normal distribution of results. (Figures 1-2) The average vocal acoustic values found in the present study for male voices producing the vowel /a/ were fo = 130.6 ± 13.65Hz, jitter = 0.046 % ± 0.184, jitter dB: 37.62 dB ± 16.66, shimmer % = 0.23%, shimmer dB = 0.23 ± 0.67 and NHR = 0.13 ± 0.010. (Figure 3) The average values found for female voices, producing the vowel /a/ were fo = 218.38 ± 26.192Hz, jitter = 0.87% ± 0.61, jitter dB: 34.82 ± 22.5, shimmer % = 2.72 ± 1.07 shimmer dB = 0.23db ± 0.67 and NHR = 0.12dB ± 0.016. (Figure 3) Fundamental frequency in females was significantly higher than their male counterparts as well as their shimmer however their NHR were slightly lower. Jitter values also showed independent variation between the two groups, Jitter % was higher in females but with a relatively lower jitter dB.

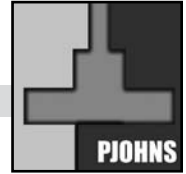
## DISCUSSION

There is a growing international trend for significant technological developments in the field of voice and speech evaluation, especially in the advancement of vocal acoustic analysis software. For this reason, standardization of normal acoustic measures is necessary due to the variation of systems protocols and software algorithms. Given the paucity of data regarding acoustic voice analysis in the Philippine literature, we decided to conceptually discuss findings obtained from the equipment used at our Voice Analysis Laboratory. As a pilot local study, we set the number of proponents in accordance to other international papers, and this may by far be the largest for this type of research compared to other studies done abroad.

Several acoustic analysis softwares have demonstrated normal and pathological voice conditions. Despite the accuracy and reliability of each machine, authors have agreed to standardize normative data individually due to a number of factors that may cause variations among each system. These possibilities include the type of programming of the acoustic analysis software, the use of recording criteria, type of microphone and other devices used in voice recording. Not only do measures vary when measured by different software; there is also a wide range of normal voices. This fact is possibly due to individual differences, since voice is a personal feature, and no voice is perfectly equal to any other.<sup>9</sup> The uniqueness of each voice also varies with race and language. These considerations led us to establish our own set of normal values for comparison data for voice analysis.

The fundamental frequency (f0) is one of the most frequently used measures by clinicians to characterize human voice and the parameter which shows uniform results among different acoustic analysis systems. The f0 is related with vocal fold length, mass and strain. Thus, lengthening the vocal folds will cause the glottic cycles to occur faster, yielding more acute resulting frequencies. Variations of this measure also result from other factors, such as different speech tasks (sustained vowels, reading, conversation, and singing) different languages and dialects, smoking, stress, dysphonia and analysis forms.<sup>5,13</sup>

Measures of the f0 using the sustained vowel /a/ in this study (Figure 3) showed a mean value of 130.62Hz ± 13.65 in the males. This value was relatively higher compared to the results obtained by Felipe *et al.*<sup>3</sup> (120Hz), Horii<sup>11</sup> (125Hz), Araujo *et al.*<sup>3</sup> (127.61Hz), Behlau and Tosi<sup>3,9</sup>(113.01Hz), and lower than those of Morente *et al.*<sup>3</sup> (139.72Hz). Measures of the f0 in the female group had a mean of 218.38Hz ± 26.19; this variation range and mean values were similar to those proposed by Araujo *et al.*<sup>3</sup> (215.42Hz) where 40 female voices issuing the vowel /a/ were evaluated using the **Analise da Voz** voice analysis software. Our values were higher than the values found by Felipe<sup>3</sup> (206Hz; CSL model 4300), Ferrand<sup>8</sup> (209.68Hz; CSL model 4300) and Finger *et al.*<sup>9</sup> (210.92Hz, Praat software) and lower than those found by Morente<sup>3</sup> (267.33Hz). This shows that our results are within the acceptable range in reference to international values and that there is a similar trend between studies



when using the  $f_0$  measure in both genders even when using other voice analysis software.

Cycle-to-cycle perturbation measures assess acoustic signal variations; they relate to how much a specific glottic vibration period is different from the ensuing period with relation to frequency (jitter) and intensity (shimmer).<sup>13</sup> Jitter, which is voice frequency cycle-to-cycle perturbation,<sup>9</sup> is an objective and reproducible measure that evaluates minor glottic pulse irregularities and may reflect hoarseness or voice noise. Jitter and shimmer have proved to be useful in the description of normal and dysphonic speakers when using sustained vowels, being respectively related to hoarseness and roughness.<sup>3,11</sup> Conversely, HNR is more sensitive to subtle differences in vocal function than is jitter according to Ferrand<sup>8</sup> after studying 42 adult women with normal voices and testing for the correlation of hoarseness and the degree of HNR. It is important to note that the results of jitter and shimmer depend on the method applied in each software and this may differ with age, sex and the vowel that is used. There are distinctive methods for extracting jitter, such as absolute jitter, relative jitter, relative average perturbation (RAP), pitch perturbation quotient (PPQ) which varies across different voice analysis softwares.<sup>9</sup>

As to the average jitter in sustained vowel /a/ among male subjects, results showed a mean value of  $0.046\% \pm 0.184$  and  $37.62\text{dB} \pm 16.66$  which was higher than the values collected by Felipe<sup>3</sup> (0.498%) and Araujo<sup>3</sup> (0.37%) but lower than the values of Horii<sup>11</sup> (0.66%). As for the females, results showed a mean Jitter  $0.87\% \pm 0.6$ , higher than the ones collected by Felipe<sup>3</sup> (0.62%), Araujo<sup>3</sup> (0.85%) and Ferrand<sup>8</sup> (0.69%).

Shimmer measures reflect the cycle to cycle amplitude variation during vibration of the vocal folds; their increase is related with a decreased or inconsistent vocal fold contact coefficient.<sup>9</sup> Different software encodes these signals in relative and absolute values however this feature may not always be present in all voice analysis programs. Furthermore, these measures may also be related with voice soprosity or noise in general.

The shimmer average for males, producing the vowel /a/ showed a relative shimmer of  $2.65\% \pm 0.76$  with an absolute shimmer at  $0.23\text{dB} \pm 0.067$ . This value was similar to the values of Felipe et al.<sup>3</sup> (0.23dB), higher than those found by Horii<sup>11</sup> (0.47 dB), but significantly lower than those of Araujo et al.<sup>3</sup> (2.37dB). Average shimmer for females producing the vowel /a/ was 0.28dB and also showed similar trends with the studies done by Felipe et al.<sup>3</sup> (0.22dB) and Finger<sup>9</sup> (0.268dB) at 2.96%. However, this was lower than the values of Araujo et al.<sup>3</sup> (2.52dB) using the **Analise da Voz** software.

A lot of controversies regarding Jitter and Shimmer parameters remain unsettled among studies and measures are not yet standardized.

The harmony-noise ratio characterizes the relationship between the two components of the acoustic wave of a sustained vowel: the periodic component, vocal fold regular sign and the additional noise coming from the vocal folds and the vocal tract.<sup>3,8</sup> A lower NHR and a higher HNR indicate superior voice quality. They reflect a general assessment of noise in a given signal. It is also influenced by age, being lower for the

elderly (from 70 to 90 years), when compared to a group of young (from 21 to 34 years) and middle age women (from 40 to 63 years).<sup>8</sup>

NHR values in our study for males and females were 0.132 and 0.117 respectively. The values for women were similar to those of Brum<sup>9</sup> (ranging from 0.03 to 0.14; mean 0.11), Schwarz<sup>9</sup> (ranging from 0.09 to 0.17; mean 0.14) and Oguz et al.<sup>9</sup> (0.157).

Despite similarities in the trends of vocal parameters in various studies, we felt the need to further explore the reference values for males as most of these papers involved female subjects.

The average vocal acoustic values found in the present study for male voices producing the vowel /a/ were  $f_0 = 130.62\text{Hz} \pm 13.65$ , jitter% =  $0.46\% \pm 0.18$ , jitter dB =  $37.62\text{dB} \pm 16.66$ , shimmer% =  $2.65\% \pm 0.76$ , shimmer dB =  $0.23\text{dB} \pm 0.067$  and NHR =  $0.132 \pm 0.009$ . The average values found for female voices, producing the vowel /a/ were  $f_0 = 218.38\text{Hz} \pm 26.19$ , jitter% =  $0.87\% \pm 0.61$ , jitter dB =  $34.82\text{dB} \pm 22.55$ , shimmer% =  $2.72\% \pm 1.07$ , shimmer dB =  $0.25\text{dB} \pm 0.105$  and NHR =  $0.117 \pm 0.016$ .

The differences in the programming of the various acoustic analysis systems, as well as the use of recording criteria, recording instrumentation such as computers, microphones and other devices individualize each of these voice acoustics systems, precluding a single normalization. Following international recommendations for individual normalization per institution, we have obtained our own values, with comparable results to other studies. This endeavor will help in our local setting establish a set of reference values for future researches in the evaluation of voice and voice related problems. In our study, the result pattern showed normal distribution of values, meeting normality of results based on Kolmogorov-Smirnov test. It is recommended that in obtaining voice samples, a strict standard procedure is followed with at least five samplings to elicit normal habitual voice and avoid false vocalization due to consciousness during the sampling. Further investigation is also suggested in areas where possibilities of interdialectic variation which may produce an effect on the outcome of the study.

#### REFERENCES

- Núñez Batalla F, Corte Santos P, Sequeiros Santiago G, Señaris Gonzáles B, Suárez Nieto C. Perceptual evaluation of dysphonia: correlation with acoustic parameters and reliability *Acta Otorrinolaringol Esp*. 2004 Jun-Jul; 55(6): 282-287.
- Toran KC, Lal BK. Objective analysis of voice in normal young adults. *Kathmandu Univ Med J*. 2009 Oct-Dec; 7(28): 374-377.
- Naufel De Felipe AC, Grillo MH, Grechi TH. Standardization of acoustic measures for normal voice patterns. *Braz J Otorhinolaryngol*. 2006 Sep-Oct; 72 (5): 659-64.
- Morris RJ, Brown WS Jr. Comparison of various automatic means for measuring mean fundamental frequency. *J Voice*. 1996 Jun; 10(2):159-65.
- Murry T, Brown WS Jr, Morris RJ. Patterns of fundamental frequency for three types of voice samples. *J Voice*. 1995;9: 282-289
- Wang CC, Huang HT. Voice acoustic analysis of normal taiwanese adults. *J Chin Med Assoc*. 2004 Apr; 67(4): 179-84.
- Tajada JD, Liesa RF, Arenas EL, Gálvez MJN, Garrido CM Gormedino PR, García AO. The effect of tobacco consumption on acoustic voice analysis. *Acta Otorrinolaringol Esp* 1999; 50(6): 448-52.
- Ferrand CT. Harmonics-to-noise ratio: an index of vocal aging. *J Voice* 2002; Dec; 16 (4):480-7.
- Finger LS, Cielo CA, Schwarz K. Acoustic vocal measures in women without voice complaints and with normal larynxes. *Braz J Otorhinolaryngol*. 2009 May-Jun; 75 (3): 432-440.
- Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*. 2001 Apr; 44(2):327-39