International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3806-3813
© International Research Publication House. https://dx.doi.org/10.37624/IJERT/13.11.2020.3806-3813

Estimation of Voice Perturbation Measures Using Signal Processing Algorithms

V. Prarthana Karunaimathi [1], D. Gladis [2] and D. Balakrishnan [3]

[1]PG & Research Department of Computer Science, Presidency College, University of Madras, Chennai, Tamil Nadu, India.

[2]Bharathi Womens' College, University of Madras, Chennai, Tamil Nadu, India [3]Department of Audiology and Speech Pathology, SRM Medical College, SRM University, Kancheepuram, Tamil Nadu, India.

ORCID: 0000-0002-9500-6602(Prarthana Karunaimathi)

Abstract

Human voice, which is the most beautiful instrument, is produced by a complicated mechanism. The quality of the voice is assessed by various techniques. Using voice analysis tools thereby providing an objective assessment of voice. These methods support the clinicians in the diagnosis of laryngeal pathologies associated with the auditory perception and also in deciding the treatment in patients. This paper aims at extraction of voice parameters from voice signals such as Fundamental Frequency (F0 Mean and Standard Deviation) Pitch Perturbation Quotient (Jiiter PPQ), Relative Average Perturbation (Jitter RAP), Amplitude Perturbation Quotient (Shimmer APQ), Pitch Variation (vF0) and Amplitude Variation (vAm) using a novel algorithm leading to an application named Ephphatha was compared with the existing software CSL (Computerized Speech Lab) using t-test analysis. The Bland-Altman plot was done showing the correlation and the limits of agreement between the measures. The outcomes got illustrate that there is no significant difference (p-value>0.05) in F0 and its variation measures. Amplitude and pitch perturbation measures are moderately (>0.01 p-value <0.05) to weakly (p-value<0.01) correlated between the programs. Thus Ephphatha would undoubtedly equip the speech pathologists and therapists with a non-invasive approach to confirm their perceptual observations.

Keywords: Acoustic Analysis; Fundamental Frequency; Pitch Perturbation; Amplitude Perturbation

I. INTRODUCTION

In human voice analysis, voice parameters play a vital role. In order to extract these parameters, the first step is to capture the voice of the subjects through a microphone and digitize them. The captured quasi-periodic speech signal consists of vital information, which can be extracted by using various algorithms. The traditional parameters that are currently in use include Fundamental Frequency (F0), Jitter, Shimmer, and Harmonics to Noise Ratio. F0 reveals the irregularity in the vocal fold, the normal falling in a particular range i.e., For an adult male, F0 ranges between 85 to 180 Hz and that for female it ranges from 165 to 255 Hz. [1] [2]. A number of researches have been carried out world-wide to extract the information in the human voice and do various analyses.

A recent study claims that the objective voice parameters are used to investigate the emotional distress of cancer patients [3]. The outcomes from this study show proof that any abnormal variation in voice fundamental frequency and amplitude depicts the presence of laryngeal pathology [4]. Most commonly the uncontrolled vibration of vocal fold increases the Jitter value which serves as a good indicator of vocal nodules and polyps [5]. This study puts evidence that the size of the vocal polyps directly affects the pitch perturbation value [6].

Another study demonstrates that the amplitude perturbation values are affected by the vocal cord lesions and reduction of glottal resistance and are exhibited by noise and breathiness

[7]. Few algorithms or applications have been developed to extract such robust parameters to improve accuracy even in the noisy acoustic environment. Such applications provide objective evidence associated with the perceptual analysis and hence can improve the treatment of patients with different types of vocal pathologies [8]. In this study, the various voice parameters are extracted from normal young adults (without any voice disorder) using a novel algorithm leading to an application called Ephphatha. The parameters extracted using Ephphatha are compared with the ones from a standard voice assessment application namely MDVP, which is proved to be a comprehensive program and widely used in the voice field for clinical and research purposes [9].

II. MATERIALS AND METHODS

The speech signals for analysis were collected from 24 female and 23 male volunteers without any voice impairment who are all in the age group of 18-25 years. The recordings were done in the speech laboratory with a microphone at a distance of 10 cm from the volunteers' lips. The task, sustained vowel phonation //a// was considered and recorded for each subject on their comfortable pitch without any interference, for at least 4 seconds. Speech signals were sampled at a rate of 44.1 kHz and 16 bits. Then the collected voice signals were analyzed with MDVP (Multi-Dimensional Voice Program) module of CSL and Ephphatha.

III. FEATURE EXTRACTION

Following are the acoustic characteristics estimated using the application Ephphatha and their description:

III.I Pitch Detection

The fundamental frequency (F0) of a voice is the rate of vibration of the vocal folds, which is perceived as pitch by the human ear. Any abnormalities in this pitch may result in a voice disorder. The Pitch Detection Algorithm (PDA) is an algorithm used to estimate the fundamental frequency of a signal with irregular periodicity. In Ephphatha, the F0 values expressed in Hz are estimated using RAPT (Robust Algorithm for Pitch Tracking) framework [10]. Mean and Standard deviation of Fundamental Frequency (MeanF0 and F0SD) gives the mean value and standard deviation of all the extracted fundamental frequency values which are expressed in equation (1) and (2) respectively.

$$\text{Mean F0} = \frac{1}{N}\sum_{i=1}^{N} F0_{(i)} \qquad (1)$$

$$\text{F0SD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F0 - F0^{(i)})^2} \qquad (2)$$

where, N is the number of extracted fundamental frequency values.

III.II Perturbation Measures

Jitter and Shimmer are the two main perturbation measures in all the acoustic analysis software, which is currently in great use in the speech labs. It allows the clinicians to use a non-invasive method to analysis the voice of the patients in order to detect whether they are pathological or not. The perturbation measures in a waveform are shown in Fig.1.

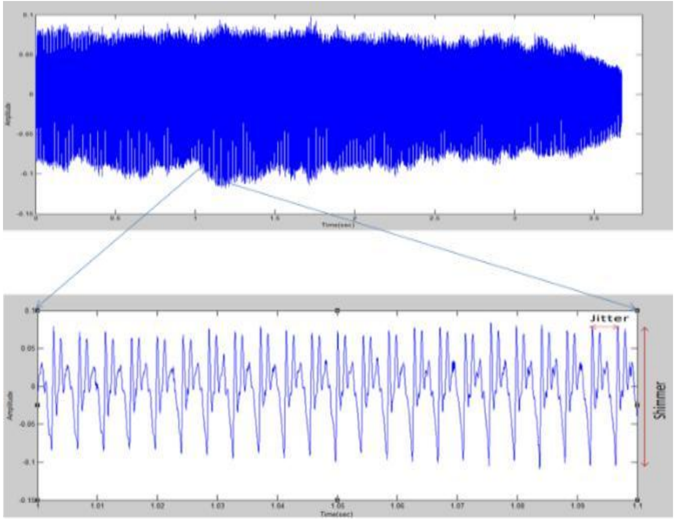


Fig .1. Sample Waveform of sustained vowel phonation //a// and its magnified version with Jitter and Shimmer representation for a healthy adult

III.II.I Relative Average Perturbation (RAP)

Jitter is the instability measure of frequency and it can be determined as the average absolute difference between successive periods which is shown in Fig. 2. Jitter RAP is used to measure the quality of voice. Voice quality is consequential in the training of vocal performers, categorically actors and singers. RAP stands for Relative Average Perturbation, which can be estimated as shown in equation [7]. It is represented in percentage as shown in equation (3).

$$\text{Jitter RAP} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|P_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} P_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} P_i} * \ 100 \qquad (3)$$

where, P_i is the duration of the i^th interval and N is the number of intervals. The interval is alternatively called as glottal period.

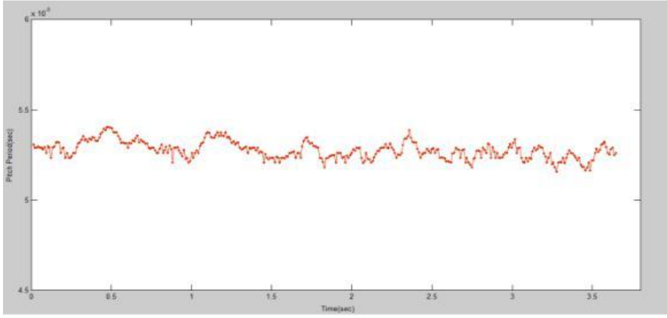


Fig. 2. Pitch Period Contour Visualization of a sample signal for normal voice in Ephphatha

For RAP, MDVP fixes 0.68% as a threshold for pathology which is comparatively higher than the other applications. The correct threshold is probably lower than this [11].

III.II.II Pitch Perturbation Quotient (PPQ)

In PPQ5, 5 represents the five-point Period Perturbation Quotien [9] which is represented in percentage. It can be estimated as shown in equation (4).

$$\text{Jitter PPQ5} = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|P_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} P_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} P_i} * 100 \qquad (4)$$

In MDVP, this parameter is named as PPQ and its value is 0.84% which is comparably higher than the other acoustic programs [12].

III.II.III Amplitude Perturbation Quotient (APQ)

Shimmer is the amplitude instability measure [13], which has been estimated in Ephphatha and it is visualized as in Fig. 3. An increase in the value of APQ explains the breathy and hoarse voice. The smoothed parameter is less sensitive to the pitch extraction error. Three smoothed factors (3, 5 and 11) are implemented in Ephphatha to extract the perturbation measures, which are discussed as follows.

1) APQ3

This is the three-point Amplitude Perturbation Quotient, which can be estimated as shown in equation (5).

$$\text{APQ3} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i-\left(\frac{1}{3}\sum_{n=i-1}^{i+1}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} *100 \qquad (5)$$

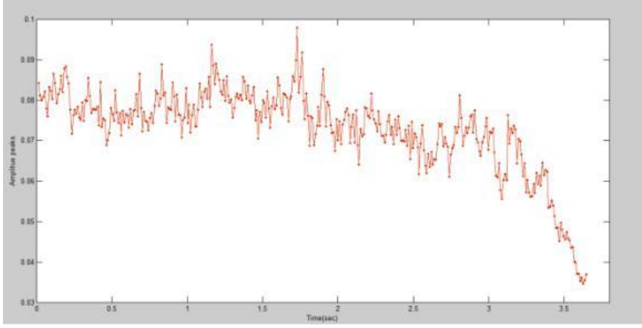


Fig. 3. Amplitude plot in Ephphatha for a sample signal of a healthy adult

2) APQ5

This is the five-point Amplitude Perturbation Quotient, which can be described as shown in (6).

$$\text{APQ5}= \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|A_i-\left(\frac{1}{5}\sum_{n=i-2}^{i+2}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} * 100 \qquad (6)$$

3)APQ11

This is the 11-point Amplitude Perturbation Quotient, is shown in equation (7). MDVP calls this parameter APQ and gives 3.070% as a t hreshold for pathology.

$$\text{APQ11}= \frac{\frac{1}{N-1}\sum_{i=5}^{N-5}\left|A_i-\left(\frac{1}{11}\sum_{n=i-5}^{i+5}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} * 100 \qquad (7)$$

III.III Variation Measures

This gives the coefficient of variation (CV) of the Fundamental Frequency (vF0) and the Amplitude (vAm). These long-term measures serve as a good indicator of vocal nodules [14]. CV elucidates the percentage of variability between the measures. It can be expressed in equation (8)

$$\text{Coefficient of Variation (CV \%)} = \frac{\sigma}{\mu} \qquad (8)$$

$$\text{where Mean } (\mu) = \frac{\sum x}{n} \qquad (9)$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum(x-\mu)^2}{n-1}} \qquad (10)$$

III.III.I Coefficient of Fundamental Frequency Variation (vF0)

It gives the relative standard deviation of F0 and it is represented in percentage. This value reflects short to long term F0 variation of the given voice sample regardless of the type of F0 variation [15]. It is formulated as the ratio of the standard deviation of the extracted F0 to the average F0 as shown in equation (11).

$$\text{vF0} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}F0^{j}-F0^{i}\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}F0^{(i)}} \qquad (11)$$

III.III.II Coefficient of Amplitude Variation (vAm)

It gives the relative standard deviation of the extracted amplitude which is represented in percentage. This value reflects short to long term amplitude variations of the given voice sample regardless of the type of amplitude variation [16]. It is formulated as the ratio of the standard deviation of the extracted amplitude to the average amplitude as shown in equation (12).

$$\text{vAm} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}A^{j}-A^{i}\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \qquad (12)$$

IV. RESULTS AND DISCUSSION

The voice attributes were estimated for a signal using Ephphatha which is developed in matlab environment and the same was done for the same segment of signal using MDVP and the corresponding variables were analyzed in both systems.

IV.I T-test

The parameter values were stored in Comma Separated Value (CSV) files, which were statistically analyzed using R Programming [17] and t-test [18] in order to obtain the P-Values, which are tabulated in Table 1. It also includes the mean and standard deviation of the parameters, t- values, degrees of freedom, and 95% confidence interval, which are always statistically parallel to the P-Values.

Table 1. Statistical results of Voice Parameters extracted using Ep[hphatha and MDVP (Age 18-25 years)

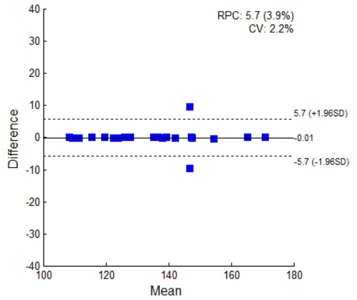
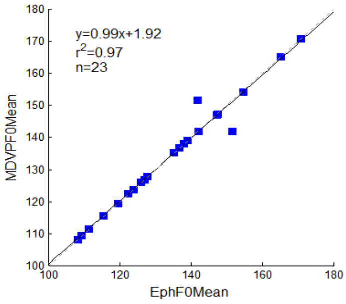
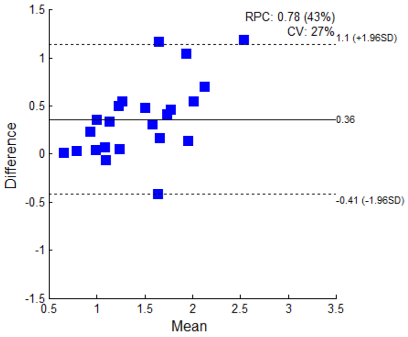
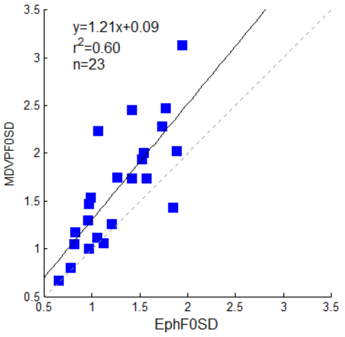
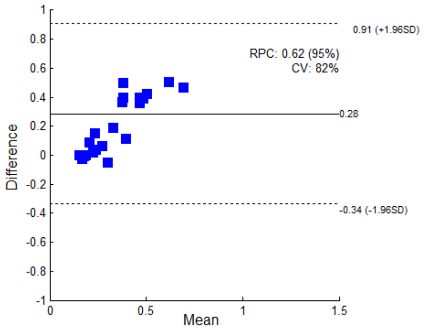
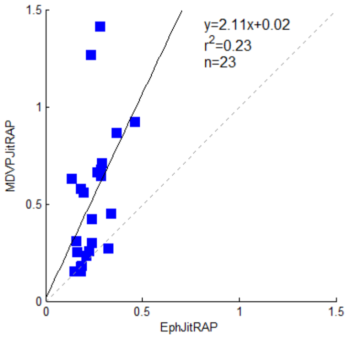
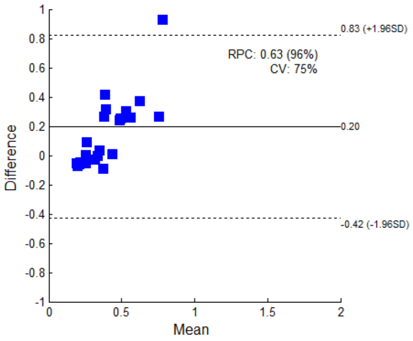
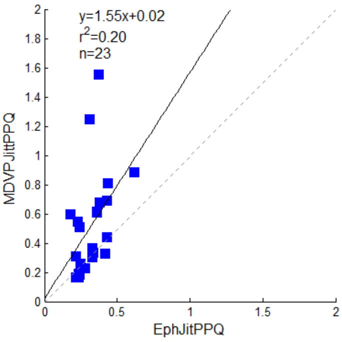
Parameters	**Program**	**Female**	**Male**	**T**	**Df**	**P-Value**	**95% confidence Interval**	
		Mean±SD	***Mean±SD***				**Lower Bound**	**Upper Bound**
F0Mean	Eph	232.30±18.04	133.39±17.84	0.0027	44	0.9978	-10.5804	10.6091
	MDVP	232.42±18.08	133.40±17.82			>0.05		
F0SD	Eph	1.8002±1.29	1.2741±0.39	2.3831	37.46	0.02235	0.0545	0.6713
	MDVP	2.8257±1.37	1.6370±0.61			>0.01		
JittRAP	Eph	0.1951±0.08	0.2427±0.08	3.8289	24.28	0.0008	0.1314	0.4383
	MDVP	0.7629±0.49	0.5275±0.35			<0.001		
JittPPQ	Eph	0.2573±0.10	0.3226±0.10	2.6354	25.69	0.0141	0.5238	0.3226
	MDVP	0.7186±0.49	0.5238±0.35			>0.01		
ShimmAPQ	Eph	3.2971±0.89	5.1951±2.73	4.3118	28.11	0.0002	-3.8749	-1.3793
	MDVP	2.9051±0.74	2.5680±1.03			<0.001		
vF0	Eph	0.7669±0.49	0.9638±0.31	2.2509	35.67	0.0307	0.0280	0.5391
	MDVP	1.2139±0.58	1.2473±0.52			>0.01		
vAm	Eph	13.4694±4.07	13.4364±6.41	1.3984	37.41	0.1702	-0.9951	5.4331
	MDVP	10.6013±2.68	11.2174±4.10			>0.05		

As seen in Table I, the outcomes got from t-test illustrates no significant difference in Amplitude Variation (vAm) (p>0.05), and that the mean of the fundamental frequency values are almost equal in both the applications since its significance level is 0.99 which is close to1. In case of F0SD, JittPPQ, and v F0, they are moderately correlated with the P-value raging from 0.01 to 0.05. The relative average perturbation and the amplitude perturbation values seem to be weakly correlated. The correlation usually gives the linear relationship between the variables but not the differences. There should be a good correlation between the variables but it is also significant to reveal the agreement between the two methods. Hence the Bland-Altman plot [19] is used to compare the newly developed method in Ephphatha with the existing method in MDVP [20].

IV.II Bland-Altman Plot

The Bland-Altman plot is otherwise called a mean-difference plot or line of agreement plot. This graphical representation is used to compare the two measurements of a variable. The x-axis denotes the mean of the two estimationss and the y-axis is the distinction between the two estimations. If the values are deviated, then it will be either above or below the zero lines. Also, the scattered points in the plot show that there is no consistent bias of one method over the other.

In the graphical representation shown in Fig. 4., the solid line represents the bias. The lines above and below the bias indicate the upper and lower lines of agreement respectively [21].

Parameters	Bland-Altman Plot	Correlation
F0Mean		
F0SD		
JittRAP		
JittPPQ		

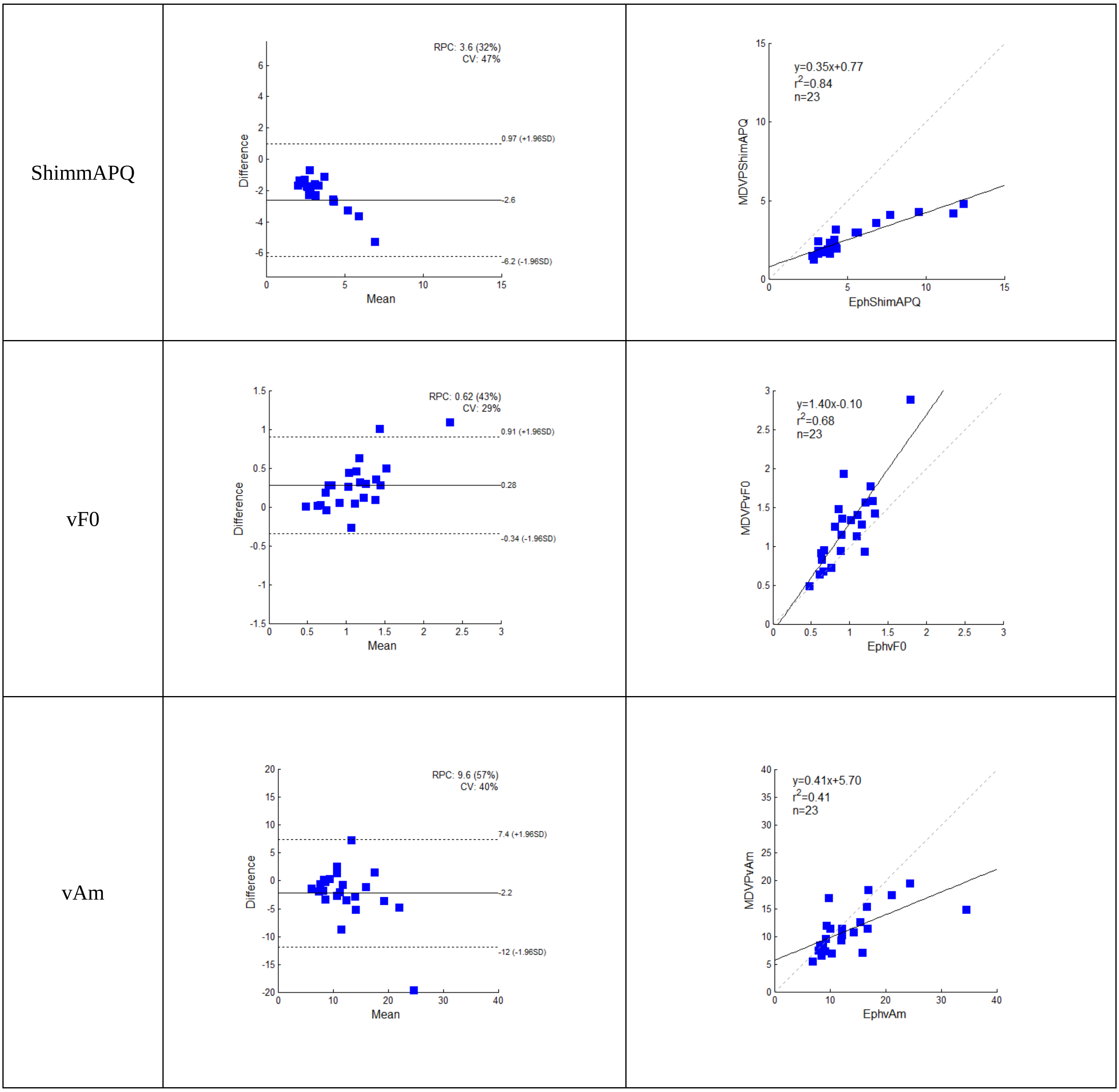


Fig. 4. Bland-Altman Agreement and Correlation Analysis of Acoustic Measurements estimated using Ephphatha and MDVP

The CV percentage shows that the fundamental frequency (F0) and its variation (F0SD and vF0) parameters are strongly correlated (CV=2.2%) in both the methods. Amplitude perturbation and variation measures are moderately correlated, since their correlation of variation is less than 50 percent. But it shows that the Jitter perturbation quotients are weakly correlated. The reproducibility Coefficient (RPC) refers to the quality of being reproducible. It is used to measure up to which extent, the resulting values of the measurement will be consistent, even if different methods are applied to the same measures under various conditions [22]. Here the variation between Ephphatha’s and MDVP's jitter measures is due to the differences between the functions in which the periods are measured. In Ephphatha, the F0 values expressed in Hz are estimated in RAPT (Robust Algorithm for Pitch Tracking) framework, which is based on NCCF [23]. . Here the window size is chosen to be in the neighborhood of the expected F0 period. The NCCF function is expressed as in equation (13).

$$\varphi_{i,k} = \frac{\sum_{j=m}^{m+n-1} s_j s_{j+k}}{\sqrt{e_m e_{m+k}}} \qquad (13)$$

where

$$e_j = \sum_{l=j}^{j+n-1} s_l^2 \qquad (14)$$

In this framework, the window size is also independent of the F0 range. It uses two versions of data signals, in which one is at the original sampled rate (44.1 kHz) and other at the reduced rate approximately 2 kHz. The peaks are refined by allowing two passes, in which the second pass searches the location which were already computed in the first pass. The improved peak is hence obtained in this framework and finally gives a candidate F0 for that frame. It solves the problem of candidate selection drawbacks occurred in the auto-correlation and cross-correlation functions by increasing the computational cost slightly. But in case of MDVP, it uses the peak picking method, which relies on the frequency response function value, and hence it is very difficult to be measured accurately. Also it cannot handle noise effectively For instance, if noise is added to a signal, the pitch perturbation in Ephphatha will give a lower value than MDVP. The threshold value for normal voice in MDVP is almost equal to the pathological level.

The limits of agreement (LoA) elucidate the range of values in which the difference between the measurements must fall with 95% probability [24]. If these limits do not exceed the maximum allowed difference between the methods, then the two methods are considered to be agreeing. It incorporates both inclination and accuracy which gives a useful measure for comparing the likely differences between singular outcomes estimated by the two techniques. It can be measured as shown in equation (15).

$$\text{LoA}= \bar{d} \pm 1.96\, S_d \qquad (15)$$

where, $\bar{d}$ and S_d are the mean and standard deviation of the differences respectively.

IV.III Coefficient of Determination (R^2)

The goodness of fit in the above regression model is shown by the Coefficient of determination (R^2). It is given as in equation (16)

$$\text{R-Squared} = 1 - \frac{SS_{regression}}{SS_{total}} \qquad (16)$$

where, $SS_{regression}$ is the sum of squares due to regression and SS_{total} is the total sum of squares.

Fig.3. shows the relationship between the voice parameters determined by Ephphatha and MDVP through the correlation graph. The parameter F0Mean value in both Eph and MDVP are perfectly and positively correlated, since their R^2 is almost 1. However, the parameters F0SD and Amplitude Perturbation Quotient, are moderately correlated with the coefficient of determination greater than 0.5. The Frequency and Amplitude Variation measures are also moderately correlated with the R-squared values close to 0.5. On the other hand, the coefficient of determination values of Pitch perturbation measures is less than 0.3, which is weakly correlated.

V. CONCLUSION

This paper focused on the analogy of voice parameters estimated from algorithms used in the application Ephphatha and the same were compared with the existing program namely MDVP. The voice signals recorded from the young healthy adults were taken into consideration for the statistical analysis. The Bland-Altman analysis was applied to ensure a thorough study of how the parameters are related between the two methods. The results reveal that there is no significant difference between the F0 values. All the other measures are moderately correlated except the pitch perturbation measures. The future heading of the study is to do a superior investigation of the algorithms in order to yield reliable measures to equip the speech pathologists and therapists with a clinical decision support system.

ACKNOWLEDGMENT

A special thanks to the Department of Audiology and Speech-Language Pathology, SRM Medical College, SRM University for supporting and providing voice data for this study.

REFERENCES

[1] Titze Ingo R., Daniel W. Martin. Principles of voice production. The Journal of the Acoustical Society of America. 1998 : vol. 104, no. 3: pp. 1148-1148.

[2] Baken Ronald J., Robert F. Orlikoff. Clinical measurement of speech and voice. Cengage Learning. 2000.

[3] Kandsberger J., Rogers S.N., Zhou Y., Humphris G. Using fundamental frequency of cancer survivors' speech to investigate emotional distress in out-patient visits. Patient education and counseling. 2016: 99(12): pp.1971-1977.

[4] Mukhopadhyay, Subhas Chandra. Advances in Biomedical Sensing, Measurements, Instrumentation and Systems. Springer. 2010.

[5] Akbari E., Seifpanahi S., Ghorbani A., Izadi F., Torabinezhad F. The effects of size and type of vocal fold polyp on some acoustic voice parameters. Iranian journal of medical sciences. 2018: 43(2): 158.

[6] Jiang J.J., Zhang Y., MacCallum J., Sprecher A., Zhou L. Objective acoustic analysis of pathological voices from patients with vocal nodules and polyps. Folia Phoniatrica et Logopaedica. 2009: 61(6): pp.342-349.

[7] Teixeira J.P. and Gonçalves A. Algorithm for jitter and shimmer measurement in pathologic voices. Procedia Computer Science. 2016: 100(100), pp.271-279.

[8] Teixeira J.P. and Fernandes P.O., Acoustic analysis of vocal dysphonia. Procedia Computer Science. 2015: 64, pp.466-473.

[9] Christmann M.K., Brancalioni A.R., Freitas C.R.D., Vargas D.Z., Keske-Soares M., Mezzomo C.L. and Mota H.B. Use of the program MDVP in different contexts: a literature review. Revista CEFAC. 2015: 17(4): pp.1341-1349.

[10] Prarthana V. Gladis D. and Dalvi U. A study of F0 estimation based on RAPT framework using sustained vowel. International Conference on Advances in Computing, Communications and Informatics, IEEE. 2015: pp. 2290-2295.

[11] Boersma P. Should jitter be measured by peak picking or by waveform matching. Folia Phoniatrica et logopaedica. 2009: 61(5): pp.305-308.

[12] Farrús M., Hernando J. and Ejarque P. Jitter and shimmer measurements for speaker recognition. Eighth annual conference of the international speech communication association. 2007.

[13] Teixeira J.P., Oliveira C., Lopes C. Vocal acoustic analysis–jitter, shimmer and hnr parameters. Procedia Technology. 2013: pp.1112-1122.

[14] Sellman J. Preliminary Evaluation of Selected Acoustic Parameters as Sensitive Indicators of Differences between Women with and without Vocal Nodules. Linguistica Uralica. 1998: Volume 3.

[15] Kahraman A. Kiliç M.A. Smoothing factor and voice perturbation measurements. B-ENT. 2011: 7(1), pp.27-30 .

[16] Geredakis A., Karala M., Ziavra N., Toki E. Preliminary Measurements of Voice Parameters using Multi Dimensional Voice Program. World Journal of Research and Review. 5(1).

[17] R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. 2013.

[18] Kim T.K.. T test as a parametric statistic. Korean journal of anesthesiology. 2015: 68(6): p.540.

[19] Bland J.M. and Altman. Statistical methods for assessing agreement between two methods of clinical measurement. The lancet. 1986: 327(8476), pp.307-310

[20] Ran Klein., Bland-Altman and Correlation Plot https://www.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot .2020.

[21] Kalra A. Decoding the Bland–Altman plot: basic review. Journal of the Practice of Cardiovascular Sciences. 2017.

[22] Bartlett J.W., Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in Obstetrics and Gynecology. The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2008: 31(4): pp.466-475 .

[23] Talkin D., Kleijn W.B. A robust algorithm for pitch tracking (RAPT). Speech coding and synthesis. 1995: 495, 1995, p.518 .

[24] Bland J.M., Altman D.G. Measuring agreement in method comparison studies. Statistical methods in medical research.1999: 8(2): pp.135-160.