Detection of Voice Disability in Children

Mullapudi Pavani, Shreyashri Athani, and Mohina Ahmadi
Electronics and Communication Engineering
PES University

***Abstract*— Voice Disability is one of the common disabilities experienced in kids. Voice being the major mode of communication it is important to rectify the voice problems at earlier stages. Painful endoscopic techniques like laryngoscopy are used by doctors to identify the voice disability. In this study, an algorithm is devised to measure the severity of voice disability in children using signal processing techniques. Spectrogram and curve fitting techniques are used to detect the voice disability. The normal and pathological curve fitted functions are passed through an adaptive signal processing system. Correlation between the normal function and tuned pathological function is obtained which is used to determine the severity of the disability.**

Keywords—Adaptive Signal Processing, Correlation Coefficients, Curve Fitting, Voice Acoustic Parameters.

I. Introduction

Voice is the sound produced by the air when it is ousted from the lungs and goes through the vocal cords. The bifold tissues within the larynx are called vocal cords and they vibrate to generate voice. The abnormal vibration of vocal cords gives rise to voice ailments.

Vocal fold vibration relies upon a few factors like the mucus existing on the vocal fold's tissue, firmness, strain, laryngeal muscles, and the opening and shutting of the vocal folds. Vibration changes starting with one kind of pathology then onto the next.

II. Related Works

Various methods have been proposed in the literature in order to detect the voice disabilities in children. In the paper by Anandthirtha et al [1], Adaptive signal processing was applied for tuning the pathological speech signals. The tuned constant values of disabled children's speech signals are compared to curve fitted speech signal constants. The deviation between them was determined using the correlation technique. The measure of deviation addresses the seriousness of speech disability in children. Adaptive signal processing techniques provided better accuracy and precision in assessing the voice disability of the subject. The limitation of this paper is the filter bank frequency bandwidth is limited which reduces the severity.

In the paper by C. Shahnaz et al [2], detection of voice disorders based on wavelet and prosody-related voice properties was presented. Statistical measures of the normalized energy contents of the Discrete Wavelet Transform coefficients and prosody-related voice properties like mean pitch, jitter and shimmer were calculated over all the voice frames. A feature vector was framed utilizing these characteristics and it was taken care of to a classifier dependent on the Euclidean distance to recognize the disabled voice. The statistical analysis on normalized energy contents of DWT coefficients was found to be useful in defining feature vectors for the detection of voice disorders. The proposed technique works well for larger data sequences while compared to the smaller data sequences.

In the paper by D.V.Borovikova et al [3], Voice Harmonization factor which is an acoustic parameter was evaluated for voice quality evaluation and pathology detection. The coefficient of voice harmonization (VHC), a spectrum parameter was developed which is able to diagnose various sorts of voice pathology and the efficacy of voice treatment. VHC is the ratio between the intensity of some high and some low harmonics. The values of the voice harmonization coefficients of the normal and pathological voices were compared and the presence of the voice pathology is determined. A relative sensitivity parameter was created to assess the received data and to identify the most sensitive voice variations of VHC. The limitation is that a number of additional studies are needed to determine the most sensitive VHC variations for various types of disorders.

In the paper by Al-Nasheri et al [4], a precise feature extraction method for voice pathology detection and classification was developed by means of autocorrelation and entropy. The features used were the maximum peak values and their matching lag values, taken from each frame of the signal through autocorrelation. Voice signal was directed to a filter bank to examine it in different frequency ranges. The output of each filter was separated into many little blocks termed frames. Peak and the lag for each frame in each specific filter's frame was extracted by applying auto correlation for one frame at a time. This was represented as a vector of features. Similarly, the entropy values for all the filter frames were represented as a feature vector. This feature vector was passed to the SVM classifier to classify the given samples into pathological and normal.

Lastly, a new set of features was represented by combining the feature extracted utilizing autocorrelation and entropy. They are sent into the SVM classifier to identify and categorize the voice disorder. The significant downside of this paper is a variation in the obtained accuracies for the same data on account of the various kinds of features used and each filter's disparate bands, as every frequency band has an alternate commitment to the recognition and characterization of disorders.

In the paper by Dina Taib et al [5], Spectrogram and formants methods were compared for feature extraction in the interest of detection of voice disabilities.In the spectrogram method, Short Time Fourier Transform was analyzed in place of a Single Fourier Transform. As, the sequence of different time varying events present in the speech waveform corresponds in the profoundly fluctuating spectral characteristics over time and they cannot be captured by a Single Fourier Transform.

In the formant frequency analysis, the formant frequencies which are the resonance frequencies of the vocal tract were investigated. The formant frequency depends on the vocal tract configuration. Better results were obtained in using spectrogram for feature extraction. The limitation in this paper is, many feature extraction methods were not included, and the data size was enormous due to which the data was sampled initially before performing the feature extraction.

In the paper by Rumana Islam et al [6], various methods of voice feature extraction methods like MFCCs, spectrograms, wavelets were analyzed. The Mel frequency cepstral coefficients (MFCCs) were determined. After windowing the voice signal, the short-time Fourier transform (STFT) was assessed. Its magnitude was then weighted by a progression of filter frequency responses whose center frequencies and bandwidth were equal to mel scale filters.

Spectrogram which is a depiction of the range of frequencies of a signal varying with time was found. Power distribution of normal voice samples with regard to time or frequency was uniform and the pathological sample was non-uniform. Subsequently, it was inferred that spectrogram could be deemed as a decent criterion to distinguish pathological voice from normal voice.

The wavelet transform is one more significant means to detect voice disabilities. Wavelets give precise data dealing with the quick variations of signals in time domain when compared to Fourier transform. Using this, there were some intermittent lines found on the graph of pathological samples which weren't found on the graph of the normal samples. Most accurate results were found for the spectrogram method to extract the voice features.

III. Proposed Methodology

"Fig. 1", illustrates the block diagram for the proposed methodology. The description of blocks is provided in this section.

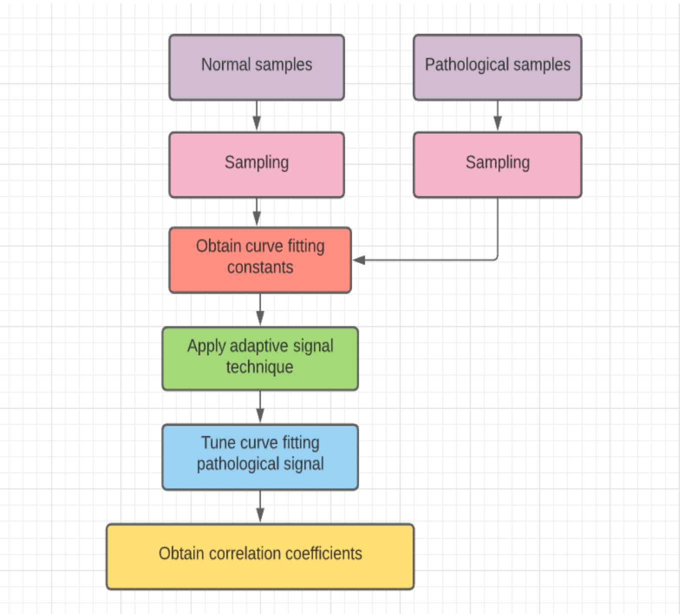

Fig. 1. Block Diagram

A. Data Collection

The voice samples are collected from the Saarbrücker Stimmdatenbank- Hauptmenü [8].This data set is in german language, audio signals which are recordings of the phrase "Good morning!,How are you?" in German, and 'a','i' and 'u' vowel sounds are extracted.

B. Detection

- Power spectral density function (PSD) is plotted which shows the strength of the changes of energy as a function of frequency. It identifies the frequencies at which the changes are strong and weak.
- Spectrogram is a visual representation of the range of frequencies of a signal as it fluctuates with time. This feature is used to analyse the data set and infer about the samples.
- Curve fitting is applied on Gaussian, Polynomial, Rational and Sum of sine functions. By considering the goodness of fit Statistics, the curve fitting method resulting in minimum error is found to be the Sum of sine function. Hence, the curve fittings for it are obtained.

C. Adaptive Filters

In time varying unknown contexts, adaptation is required. Adaptive filters are used in applications where adaptation of a process or a low processing delay is required. An adaptive filter is a digital filter with self-adjusting characteristics. The typical examples of adaptative filtering are echo cancellation, channel equalization, speech coding etc.

The input is passed into a filter and error signal is used as feedback to improve its transfer function. Adaptive algorithm updates the weights of filter coefficients to minimize the error. This is done till these weights reach an optimal value. Filtering and Weight adjustment are the two stages in adaptive filters.

[New weight=Old weight + Adaptation Gain x Estimated Error]

Convergence speed is the rate at which the filter meets the desired state. It relies on different factors like the nature of the input signal, type of the adaptive filter algorithm and its step size. The choice of filter algorithm will change based on the factors like convergence performance that are needed for application, filter stability in the environment.

LMS (Least Mean Square) filters are mostly used adaptive filters. The algorithm is simple and efficient. These filters are robust to noise. Adaptation depends on the gradient-based methodology that updates filter weights to converge to the desired filter weights. Based on the error at the current time, this class of algorithms are adapted.

LMS algorithm is sensitive to the scaling of its input which makes the determination of a learning rate that ensures its stability to be troublesome. Hence, *Normalized LMS algorithm* is introduced. It is a variation of the LMS algorithm that tackles this issue by normalizing with the power of input. Convergence speed of NLMS filter is faster than LMS Filter. Therefore, NLMS Filter is applied in this work.

D. Estimation

- A procedural algorithm of adaptive signal processing is written, such that it takes input as voice signal samples and takes a curve fitting equation in the feedback loop to detect and correct error. The pathological sample values are adapted to the normal samples.
- Curve fitting equation of the normal subject is correlated with the tuned pathological sample which was obtained after the adaptive process.
- The correlation coefficient is calculated for ‘a’,’i’,’u’ sounds and their average value gives the measure of severity of voice disability in the child.

E. Voice Acoustic Parameters

Studies show that pathological voices can be identified by the analyzing the acoustic signal parameters [7].

1) Jitter(local): It is the ratio of average absolute difference between consecutive periods and the average period.Jitteriness in voice can be measured in various ways.The most generic way is to find the variation or unsteadiness between pack delay, referring to variation in frequency due to congestion caused from voice disability.The value of jitteriness varies highly from one voice disability to another.It can even vary based of the gender.

2) Shimmer(local): It is the ratio of the average absolute difference between the amplitudes of consecutive periods and the average amplitude.Very similar to jitter, shimmer in voice can also be measured in various ways along with different parameters, but the major difference here is unsteadiness of the voice signal is caused by variations in the signal amplitude.

3) HNR Ratio:It measures the amount of additive noise in the voice signal.It acts as one of the major vocal acoustic sensitive indexes.Anyhow, this measure is reliant with the kind of window used and its length.It is likewise considered as a parameter to evaluate the degree of hoarseness in the voice.It shows the strength of harmonic (periodic) levels over noise (aperiodic) levels measured in dB.

*a) Hoarseness:*When hoarse, the voice sounds rough,breathy, or strained.There might be changes in the volume or pitch.Hoarseness is a symptom and not a disease.It is an overall term that depicts strange voice changes. The changes in voice are generally because of the issues related to vocal folds.

IV. Experiments and Results

The procedure proposed in methodology is performed and the results are explicated in this section.

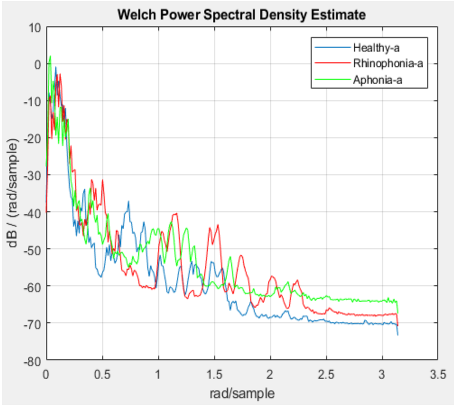

Fig. 2. PSD estimate of Normal vs Pathological samples for ‘a’ sound

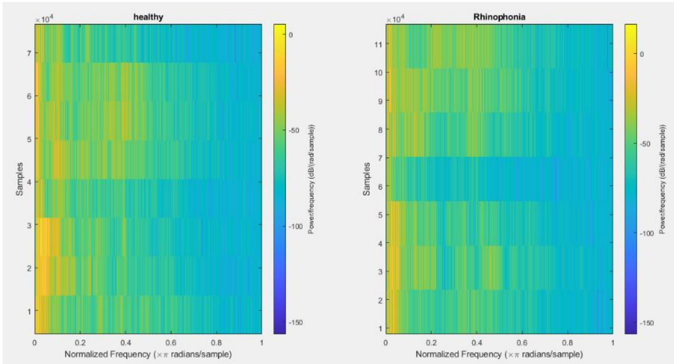

Fig. 3. Spectrogram of Normal sample and a Pathological sample

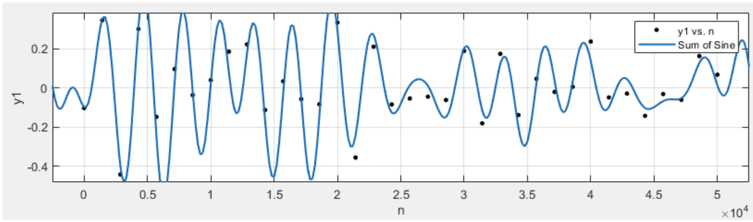

Fig. 4. Curve fitting on Sum of Sine function

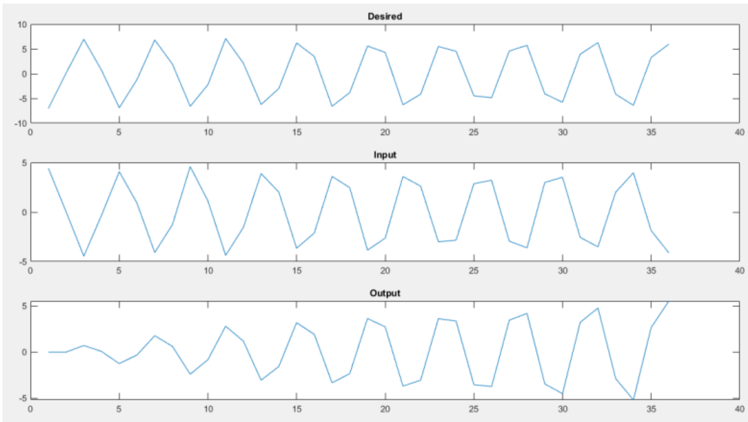

Fig. 5. Plot of the desired signal, input signal and, ouput of the filter

TABLE I. EVALUATED CORRELATION COEFFICIENTS OF SAMPLES

S.No	Gender	Type	A	I	U	R
1	Female	Healthy	0.9035	0.5617	0.4149	0.696
2	Female	Rhinophonia Aperta	0.0513	0.2335	0.2695	0.186
3	Female	Aphonia	0.1930	0.2848	0.0776	0.186
4	Female	Laryngitis	0.3527	0.0701	0.1072	0.170
5	Female	Velopharyng -oplasty	0.0608	0.3281	0.1234	0.174
6	Male	Healthy	0.2409	0.7794	0.4647	0.535
7	Male	Phonation Nodules	0.2133	0.4158	0.1826	0.274
8	Male	Rhinophonia Aperta	0.2757	0.3678	0.0505	0.235
9	Male	Juvenile Dysphonia	0.0454	0.1444	0.0085	0.066
10	Male	Laryngitis	0.0745	0.0399	0.1588	0.091
11	Male	Recurrent Palsy	0.1585	0.1424	0.5649	0.304

TABLE II. VOICE ACOUSTIC PARAMETERS OF SAMPLES

S. No	Type	R	Sound	Jitter (%)	Shimmer (%)	HNR (dB)
1	Healthy	0.696	a	0.315	2.286	25.634
			i	0.176	1.582	26.855
			u	0.238	2.546	28.505
2	Rhinophonia Aperta	0.186	a	0.629	4.403	21.417
			i	0.318	0.591	26.407
			u	0.280	2.029	27.444
3	Aphonia	0.186	a	0.314	2.698	24.617
			i	0.130	1.190	29.739
			u	0.245	1.602	30.675
4	Laryngitis	0.170	a	0.350	3.586	19.170
			i	0.443	2.384	20.244
			u	0.678	2.363	24.260
5	Velopharyng oplasty	0.174	a	0.388	4.358	17.208
			i	0.307	2.307	20.803
			u	0.616	3.941	22.296
6	Healthy	0.535	a	0.271	2.440	21.328
			i	0.440	1.714	28.559
			u	0.193	2.334	28.026
7	Phonation Nodules	0.274	a	0.881	6.891	12.612
			i	0.946	3.620	18.922
			u	0.762	4.183	22.407
8	Rhinophonia Aperta	0.235	a	0.463	2.356	23.050
			i	0.385	1.076	30.476
			u	0.205	0.802	34.933
9	Juvenile Dysphonia	0.066	a	0.218	1.723	28.045
			i	0.270	0.925	24.581
			u	0.521	1.796	27.711
10	Laryngitis	0.091	a	0.559	3.246	18.015
			i	0.540	1.977	26.347
			u	0.168	0.963	35.288
11	Recurrent Palsy	0.304	a	0.565	7.846	18.800
			i	0.353	3.597	25.019
			u	0.305	2.719	23.292

PRAAT Software [9] is used to extract the voice parameters like Jitter, Shimmer and HNR for the sound samples.

According to the PRAAT program, “the threshold of pathology is <=1.040% for jitter and <=3.810% for shimmer”.Any percentage score above these exhibit a deviation from expected typical functioning which can be considered as a sign of potential pathology. HNR below 20 dB is viewed as a measure of perceptible hoarseness.

In Table II, Samples-2,5,7,8,11 have a greater shimmer value than the threshold value and Samples-4,5,7,10,11 have an HNR ratio which is less than 20 dB which indicates that, these voices are hoarse.

TABLE III. THRESHOLD VALUES OF VOICE ACOUSTIC PARAMETERS

JITTER	<=1.040%
SHIMMER	<=3.810%
HNR	>20 dB

Based on the above findings, a range is defined which gives the measure of severity of the voice disability.

TABLE IV. RANGES OF SEVERITY OF THE DISABILITY

Correlation Coefficient	Severity
$0 \leq R < 0.2$	High
$0.2 \leq R < 0.5$	Moderate
$0.5 \leq R \leq 1$	Low or Nil

The ranges can change based on the quantity and quality of healthy samples considered as desired samples. More the number of samples, better will be the results.

TABLE V. SEVERITY OF THE DISABILITY IN THE TEST SAMPLES

S.No	Age(y)	Gender	R	Severity
1	4	Male	0.123	High
2	5	Male	0.459	Moderate
3	11	Male	0.235	Moderate
4	5	Female	0.141	High
5	10	Female	0.195	High
6	11	Female	0.250	Moderate
7	13	Female	0.284	Moderate

V. CONCLUSION AND FUTURE WORK

This work presents a method to detect voice disability and estimate the level of severity of the disability in children. This is implemented using curve fitting techniques, concepts of adaptive signal processing and by analyzing the impact of voice acoustic parameters on the disabilities.

More numbers of samples can be included to get better accuracy. The methods used to collect the voice samples affect the parameters and the results obtained. Hence, the voice samples should be recorded carefully using specific instruments in noiseless environments. This algorithm can be further improved to detect the type of disability using more parameters.

ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Shikha Tripathi, Department of Electronics and Communication Engineering, PES University for her continuous guidance, assistance, and encouragement throughout this work.

We are deeply grateful to PES University for providing various opportunities and enlightenment.

REFERENCES

[1] Anandthirtha. B. Gudi, H.K. Shreedhar and H. C. Nagaraj,”Signal Processing Techniques to Estimate the Speech Disability in Children”, IACSIT International Journal of Engineering and Technology, Vol.2, No.2, April 2010 ISSN: 1793-8236

[2] C. Shahnaz,S.A.Fattah,U.Mahbub,”Detection of Voice Disorders based on Wavelet and Prosody-related Properties”,IEEE International Symposium circuits and systems(ISCAS),August 2012,doi:10.1109/ISCAS.2012.6271403

[3] D. V. Borovikova, V. K. Makukha and A. E. Tsvetkov, "Comparative analysis of different Voice Coefficient variations for children's voice disorders detection," 2016 13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), Novosibirsk, 2016, pp. 432-433, doi: 10.1109/APEIE.2016.7802193

[4] A. Al-Nasheri et al., "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions," in IEEE Access, vol. 6, pp. 6961-6974, 2018, doi: 10.1109/ACCESS.2017.2696056

[5] Dina Taib, Mohammad Tariq, Rumana ,”Voice Feature Analysis for Early Detection of Voice Disability in Children”,IEEE International Symposium on Signal Processing and Information Technology ,December 2018, doi:10.1109/ISSPIT.2018.8642783

[6] Rumana Islam, Mohammed Tarique ,Esam Abdel-Raheem,”A Survey on Signal Processing Based Pathological Voice Detection Techniques”,IEEE Access,April 2020,DOI:10.1109/ACCESS.2020.2985280

[7] João Paulo Teixeira*, Carla Oliveira, Carla Lope, “Vocal Acoustic Analysis - Jitter, Shimmer and HNR Parameters” CENTERIS 2013 - Conference on ENTERprise Information Systems / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies.

[8] W. J. Barry and M. P¨utzer, “Saarbr¨ucken Voice Database”, Institute of Phonetics, Univ. ofSaarland.

[9] Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* **5:9/10**, 341-345.