Speech Signal Analysis and Classification of Dominant Parameter for Pathological Voices

Christina Subiksha W, Nandhini A, Bharath K P, Mahalti Mohammed Sohail, Rajesh Kumar M

Abstract*: *The primary objective of the project is to analyze speech signals by determining the important parameters that affect the voice of an individual which leads to various voice disorders. The analysis is carried out based on the individual's age and gender with the help of the pattern recognized from each sample and the value of each parameter is compared with the nominal values of the healthy person with respect to their age and gender using the Praat software. The secondary objective is the classification of the voice signal into normal and abnormal voice samples using the machine learning software Konstanz Information Miner (KNIME).

Keywords*: *Harmonics-to-noise ratio (HNR), Jitter, Konstanz Information Miner (KNIME), Praat, Shimmer.

I. INTRODUCTION

Voice disorders are medical conditions associated with irregular pitch, nature of the sound given by the larynx and thereby changing speech or sound production thus leads to difficulty in communicating with others. The voice disorders depends on various factors like age, gender, cultural background or the geographic location of an individual. Most of the voice disorders are curable at an early stage with proper diagnosis and treatment therefore it is extremely essential for pathologists and speech therapists to have more knowledge in recognizing the speech patterns between a normal and abnormal patients. Elizabeth S. Hasseltine et al. [1] have stated that voice disorders that disturb children and adults have various reasons as for various age gatherings. Factors related with voice issue varies from medical to stress. This research was focused to consider the various pathological reasons prompting to abnormal voice changes. Mansi Kumbhakarn et al. [2] have stated that features like formant frequencies, pitch, intensity and so on are benefited for recognition of emotion.

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author

Christina Subiksha W, School of Electronics Engineering, VIT University, Vellore, India. Email: teenawaltz@gmail.com

Nandhini A, School of Electronics Engineering, VIT University, Vellore, India. Email: 18nandhiniakilan@gmail.com

Bharath K P, School of Electronics Engineering, VIT University, Vellore, India. Email: bharathkp25@gmail.com

Mahalti Mohammed Sohail, School of Electronics Engineering, VIT University, Vellore, India. Email: msohail17925@gmail.com

Rajesh Kumar M*, School of Electronics Engineering, VIT University, Vellore, India. Email: mrajeshkumar@vit.ac.in

All these features are valued together to eventually categorize four emotions especially sad, angry, neutral and happy. Betul Erdogdu Sakar et al. [3] have stated that an advanced concern in speech pattern inquiry utilization of Parkinsonism for developing predictive telediagnosis and telemonitoring designs. For that reason, they have gathered a large collection of voice samples, comprising vowels, words and sentences gathered from a collection of speaking activity for Parkinson's patients. Carlos Busso et al. [4] have analyzed the advantages and the disadvantages of schemes based on facial aspects or data related to sound. It also deliberates two concepts benefited to combine these two methods: decision level and feature level integration. Utilizing a database recorded from an actress, four emotions were divided: happiness, sadness, neutral state, and anger. By the benefit of indicators on her face, accurate facial gestures were taken with gesture capture, in combination with concurrent speech recordings. The conclusions admit that the scheme based on facial aspect gave improved performance than the scheme based on data related to sound for the feelings considered.

Minu George Thoppil et al. [5] have determined the harshness of dysarthria by analyzing speech using Praat and MATLAB toolkit. Tan Tze Ern Shannon et al. [6] have analyzed quantitatively to determine harshness of depression using features of speech. Haritha C. K. et al. [7] have analyzed the parameters of voice using Praat software in 20-25 years adults and established norms. Magdalena Majdak et al. [8] have made a study to regulate the acoustic features of the influence of a voice training intervention using Praat and MATLAB. Bahman Gorjian et al. [9] have made a study to analyze the performance of Praat in facilitating students to gain prosodic features of English. Lamia Bouafif et al. [10] have established an interface using MATLAB toolbox which is used for speech coding, recognition and denoising and compared with other softwares like Phonedit, speech analyzer and Praat. Sonu et al. [11] have analyzed the voice parameters for identifying disease using feature extraction (Mel Frequency Cepstral Coefficient) and feature matching (Dynamic Time Warping). In our work, the voice sample analysis was carried out using the Praat software and classification of voice signal using KNIME.

II. DESIGN APPROACH

A. Speech Signal Analysis

Speech signal is given as an input. The sound samples were collected from 50 subjects. The subjects were made to tell a sentence of maximum 5 to 10 seconds duration.

The subjects were made to say a common sentence, "Hello… hello … hello".

The sound was recorded in a noise proof space to avoid any noise in the speech signal. The speech samples were collected from 50 subjects that include 25 male subjects and female subjects. The subjects were chosen based on their age and gender. The subjects were classified into five groups in an ascending order. The age groups are 1–15 years, 16–25 years, 26–40 years, 41–60 years and above 60 years. The subjects were also made to pronounce syllables like "ba", "ka", "ma", "pa" and "a" to get more clear spectrum analysis on particular syllables. The sound recordings were taken as an input in the Praat software and analyzed with acoustic and spectrum analysis of the spectrum window in Praat. The software also yields a voice report of the sound sample fed in the software. The report explains about all the parameters and the corresponding values of the sound sample. Once the sound sample is read and the sound sample is satisfied enough to analyze the sound is added to the object window. The sampling frequency is set as shown in Fig. 1. The default sampling rate is 22000 Hz although the sampling rate depends on the system's disk space. From the pulse menu, the voice report for the speech signal can be derived. This voice report provides all the measurements of the voice signal as shown in Fig. 2.

- *Jitter (local):* The average absolute difference between consecutive periods / the average period. (Threshold value - 1.040%)
- *Jitter (local, absolute):* The average absolute difference between consecutive periods. (Threshold value - 83.200 seconds)
- *Jitter (rap):* The Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbours / the average period. (Threshold value - 0.680%)
- *Jitter (ppq5):* The five-point Period Perturbation Quotient, the average absolute difference between a period and the average of it and its four closest neighbours / the average period. (Threshold value - 0.840%)
- *Jitter (ddp):* This is the average absolute difference between consecutive differences between consecutive periods / the average period.
- *Shimmer (local):* The average absolute difference between the amplitudes of consecutive periods / the average amplitude. (Threshold value - 3.810%)
- *Shimmer (local, dB):* The average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods x 20. (Threshold value - 0.350 dB)
- *Shimmer (apq3):* The three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours / the average amplitude.
- *Shimmer (apq5):* The five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours / the average amplitude.
- *Shimmer (apq11):* The 11-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbours / the average amplitude. (Threshold value - 3.070%)

The voice is further analyzed to extract the spectrum of the voice samples. Fig. 3 shows the spectrum of voice sample. The spectrum has four analyses in its spectrogram. The red plot in the spectrum indicates the formant frequencies. The number of formant frequencies is 5. Thus five plots can be seen from Fig. 4.

The blue plot in the spectrum indicates the pitch of the sound sample. It is shown in Fig. 5. The yellow plot in the spectrum indicates the intensity of the sound sample. It is shown in Fig. 6. The blue lines in the waveform are the pulses of the speech signal. It is shown in Fig. 7.

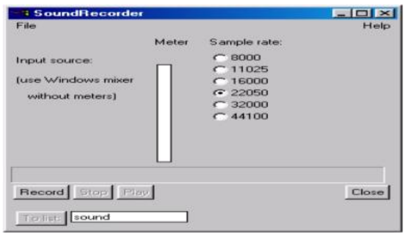
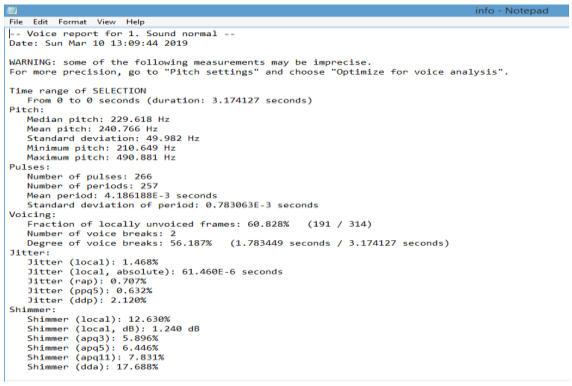


Fig. 1. Setting sample frequency.



```
info - Notepad
File Edit Format View Help
-- Voice report for 1. Sound normal --
Date: Sun Mar 10 13:09:44 2019

WARNING: some of the following measurements may be imprecise.
For more precision, go to "Pitch settings" and choose "Optimize for voice analysis".

Time range of SELECTION
   From 0 to 0 seconds (duration: 3.174127 seconds)
Pitch:
   Median pitch: 229.618 Hz
   Mean pitch: 240.766 Hz
   Standard deviation: 49.982 Hz
   Minimum pitch: 210.649 Hz
   Maximum pitch: 490.881 Hz
Pulses:
   Number of pulses: 266
   Number of periods: 257
   Mean period: 4.186188E-3 seconds
   Standard deviation of period: 0.783063E-3 seconds
Voicing:
   Fraction of locally unvoiced frames: 60.828%   (191 / 314)
   Number of voice breaks: 2
   Degree of voice breaks: 56.187%   (1.783449 seconds / 3.174127 seconds)
Jitter:
   Jitter (local): 1.468%
   Jitter (local, absolute): 61.460E-6 seconds
   Jitter (rap): 0.707%
   Jitter (ppq5): 0.632%
   Jitter (ddp): 2.120%
Shimmer:
   Shimmer (local): 12.630%
   Shimmer (local, dB): 1.240 dB
   Shimmer (apq3): 5.896%
   Shimmer (apq5): 6.446%
   Shimmer (apq11): 7.831%
   Shimmer (dda): 17.688%
```

Fig. 2. Voice report of a voice sample.

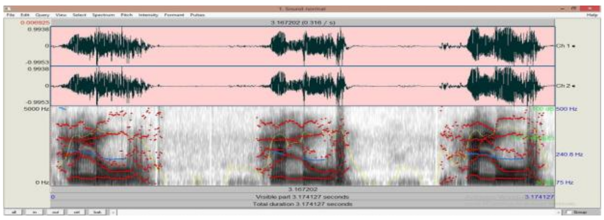


Fig. 3. Spectrum of voice sample in Praat software.

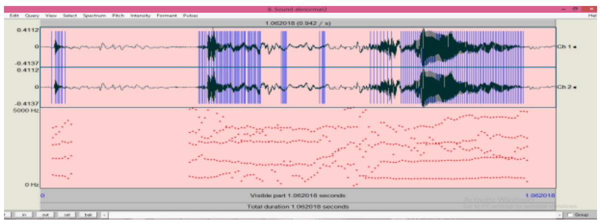


Fig. 4. Spectrum showing the five formant frequencies.

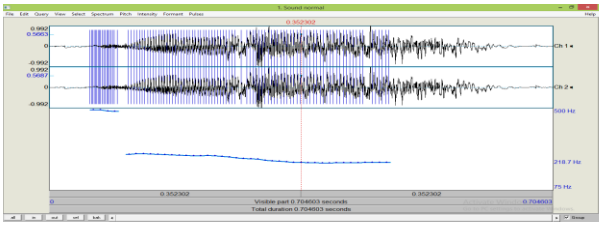


Fig. 5. Spectrum showing the pitch contour.

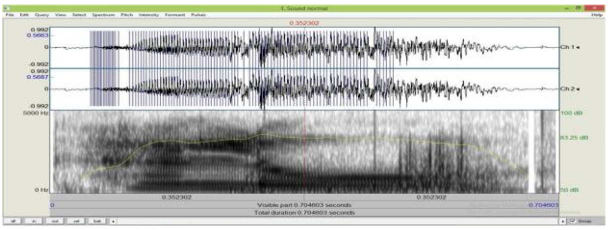


Fig. 6. Spectrum showing the intensity contour.

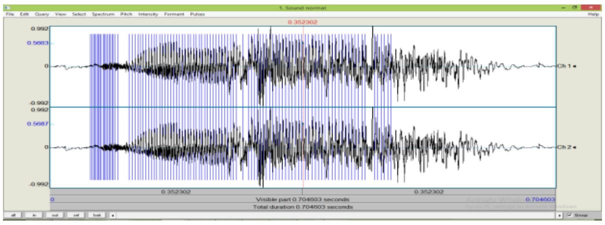


Fig. 7. Spectrum showing the pulses in the speech signal waveform.

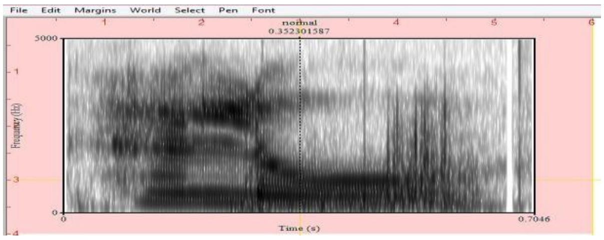


Fig. 8. Spectrum of a voice signal.

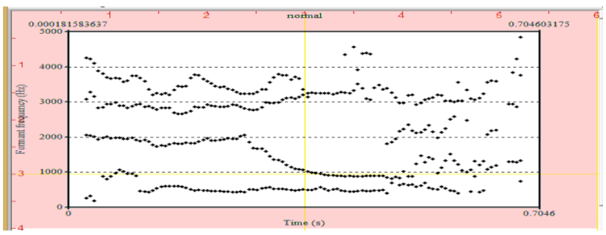


Fig. 9. Plot of Formant frequency.

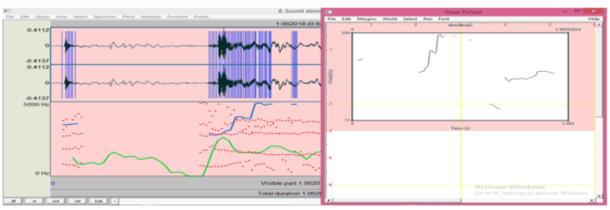


Fig. 10. Plot of Pitch contour.

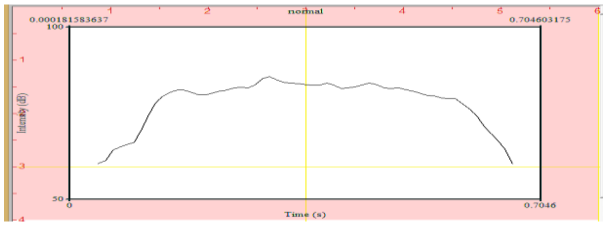


Fig. 11. Plot of Intensity contour.

B. Spectrum Analysis

Fig. 8 shows the spectrogram of the voice signal. In this spectrogram, the dark shaded pattern depicts the pattern of waveform of the voice signal. Fig. 9 is the formant analysis of voice signal spectrum. Fig. 10 shows the pitch contour analysis of the voice signal spectrum. Fig. 11 shows the plot of intensity contour of the voice signal spectrum. Once the acoustic parametric analysis and spectrogram analysis are completed. The next step of analysis is classification of speech signals between normal and abnormal voice samples.

C. Parametric Analysis

The following nominal threshold values of each parameter are decided based on the speech pathologist of the speech therapies of the hospitals.

- *Normal pitch range:* The normal pitch range of a pathological voice is 60 – 400 Hz. The pitch floor is 75 Hz.
  - *Female:* The pitch range for female is however much larger than that of males. The normal pitch range for a female voice is 100 – 525 Hz.
  - *Male:* The male voices have lesser pitch range than that of females. The normal pitch range of male voices is 65 – 260 Hz.
- *Jitter range:* Jitter is the change in periodicity of signal. The normal jitter range for pathological voices is 1.080%. The jitter range greater than the optimum range is regarded as abnormal.
- *Shimmer range:* Shimmer is a measure of amplitude instability in the voice. The normal shimmer range for pathological voices is 3.840%. The shimmer range greater than the optimum range is regarded as abnormal.
- *Harmonics-to-noise ratio (HNR) range:* A Harmonicity object describes the level of acoustic periodicity. The optimum HNR value for normal healthy voices is 20 dB. The HNR values lesser than that of the optimum value is considered to be abnormal.
- *Voice breaks:* The normative value for the fraction of unvoiced frames is 0.

D. Voice Signal Classification

The next step in the analysis is classification. For classification of voice signal to be normal or abnormal, the Knime software is used. The workflow model is shown in Fig. 12. The data is acquired and selected using the file reader node and used for further analysis. It is shown in Fig. 13.

After acquiring the data it is checked for missing values then passed onto the numeric binner, here we bin the data into two parts, normal and abnormal as you can see from the excel sheet values till 5 are abnormal. Values after 5 are normal hence using this we created two bins. It is shown in Fig. 14 and Fig. 15. The data is passed through the decision tree learner to train the working model to give the appropriate results. The partitioning allows only a certain amount of data into decision tree learner and predictor. It is shown in Fig. 16. The Decision tree predictor gives the most suitable results to classify data depending upon the amount of input. It is shown in Fig. 17.

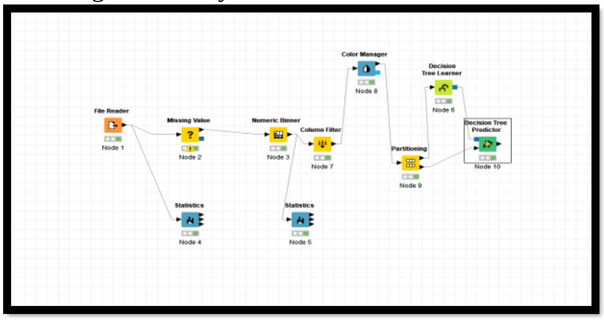


Fig. 12. Workflow model of the signal parameter classification.

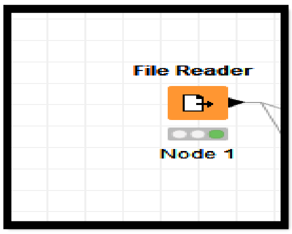


Fig. 13. The file reader component.

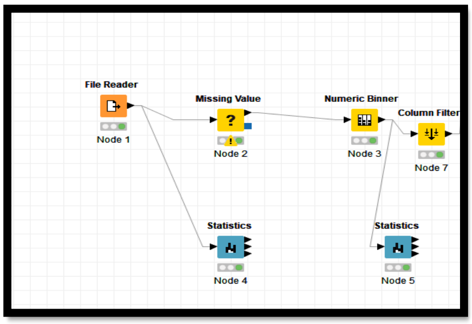


Fig. 14. Data storing, correction and allocation.

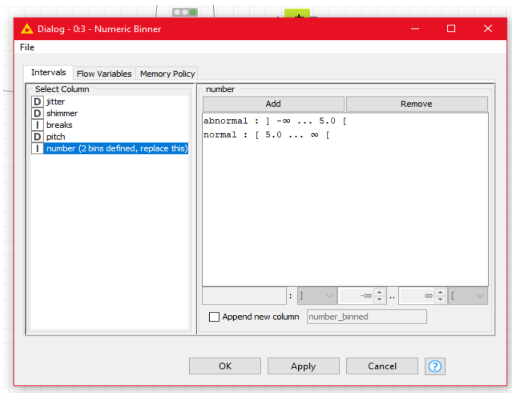


Fig. 15. Data categorized in numeric binner.

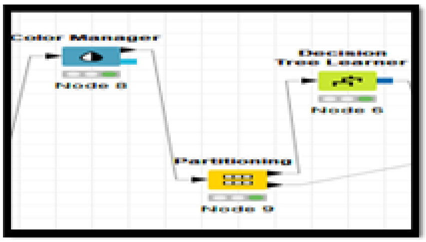


Fig. 16. Data training and input data.

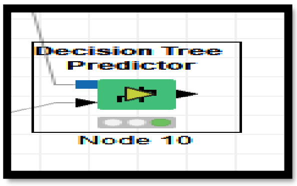


Fig. 17. Decision tree predictor.

III. RESULTS

A. Parametric Analysis

On analysis of 50 pathological voice samples the following results on their average values is yielded. From Table-I, the jitter, the shimmer and the HNR values show much variation than the nominal values of the parameters. The jitter values keep increasing with increase in age. The shimmer values are higher than the nominal value of shimmer range. The HNR values are much lesser than the normal HNR value of 20 dB. From Table-I, we have analyzed the pathological voice samples to derive the parametric analysis of the samples and perform the classification of voice signals.

B. Spectrum Analysis

1) Spectrogram analysis

Fig. 18 shows the spectrogram of a male voice sample of the age 15 – 25 years have less number of blue lines that represents the pulses in the voice. Fig. 19 shows the spectrogram of a male voice sample of the age of above 65 years have more number of blue lines that represents the pulses of the voice. This indicates that the older age voice is not as stable as the younger age voice and as a result the younger age voice has more pulses in its spectrogram.

2) Waveform analysis

Fig. 20 shows the waveform of a normal voice sample which has no voice breaks. Fig. 21 shows the waveform of an abnormal voice sample which has long breaks.

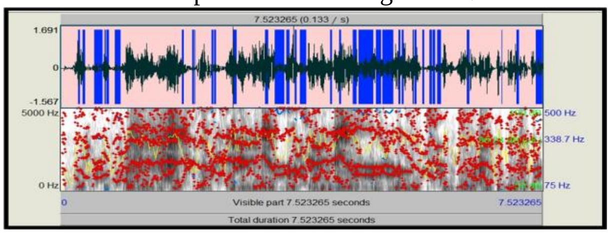


Fig. 18. Spectrogram of male voice of (15-25) years age.

Table-I Analysis of average sample for each age group

Features	Age group									
	0-15		*15-25*		*26-40*		*41-60*		*>60*	
	Male	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
Pitch (Hz) (Range: 60-400 Hz)	114.23	107.85	181.89	176.543	100.987	102.364	68.991	89.377	103.09	105.77
Jitter (Normal range: <1%)	2.84%	2.124%	2.61%	4.71%	4.23%	5.69%	5.61%	5.50%	6.37%	7.75%
Shimmer (Normal range: <3%)	20.36%	18.50%	20.60%	21.71%	18.13%	20.18%	27.34%	17.61%	19.61%	18.66%
HNR (Normal range: >20dB)	5.968	7.987	3.764	4.765	6.981	5.743	5.072	3.544	4.327	3.989

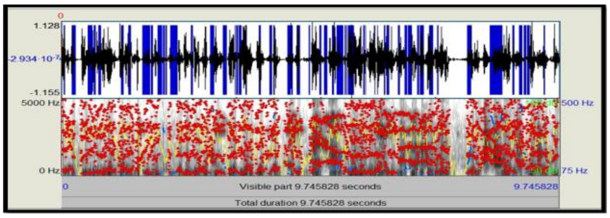


Fig. 19. Spectrogram of male voice of above 65 years age.

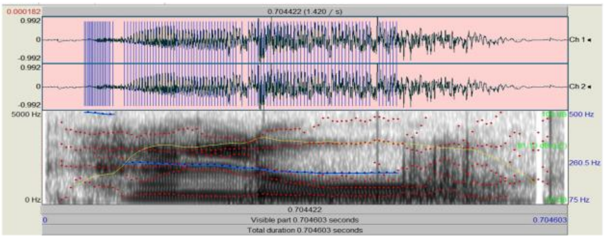


Fig. 20. Waveform of a normal voice.

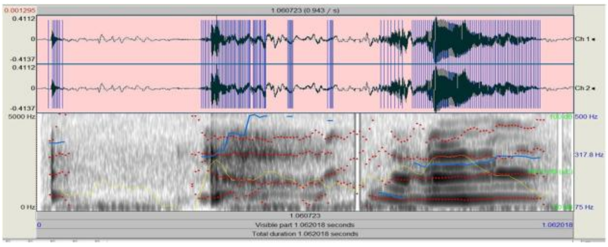


Fig. 21. Waveform of an abnormal voice.

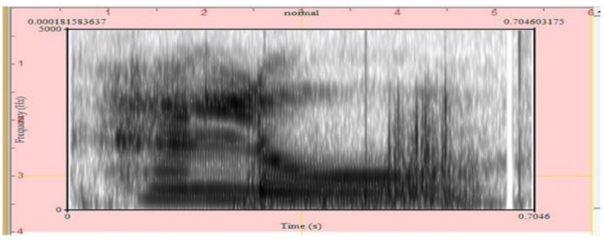


Fig. 22. Spectrum of a normal voice.

3) Spectrum analysis

Fig. 22 and Fig. 23 show the spectrum of a normal voice sample which has a regular pattern and the spectrum of an abnormal voice sample which has an irregular pattern and has breaks between voiced spectrum respectively.

4) Pitch contour analysis

Fig. 24 and Fig. 25 show the pitch contour of a normal voice which has a linear plot and the pitch contour plot of an abnormal voice has a broken plot respectively.

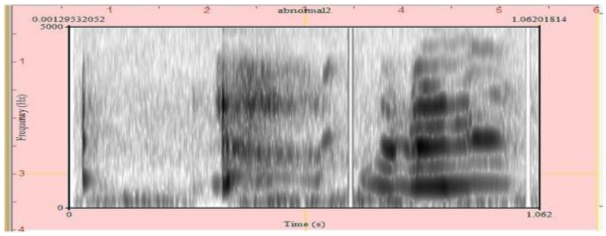


Fig. 23. Spectrum of an abnormal voice.

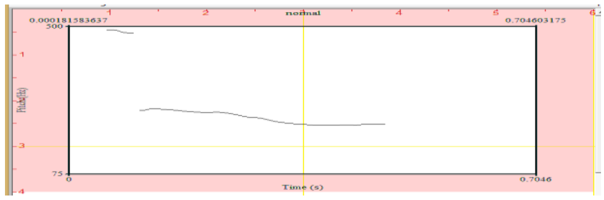


Fig. 24. Pitch contour of a normal voice.

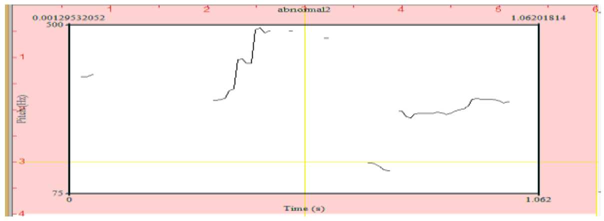


Fig. 25. Pitch contour of an abnormal voice.

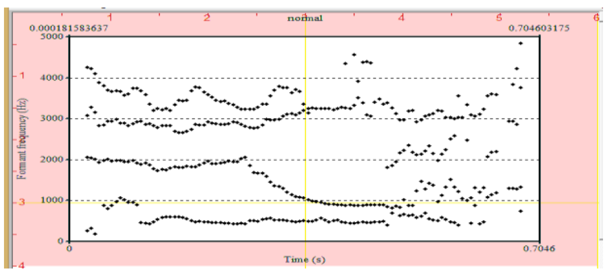


Fig. 26. Formant frequencies of a normal voice.

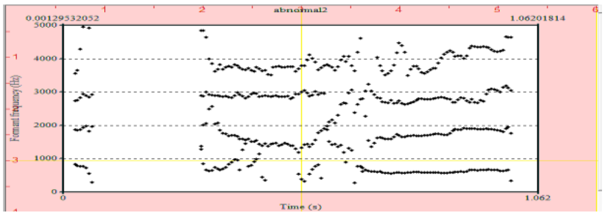


Fig. 27. Formant frequencies of an abnormal voice.

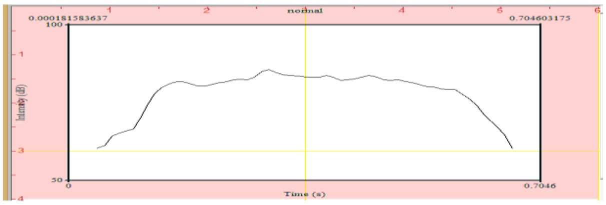


Fig. 28. Intensity spectrum of normal voice.

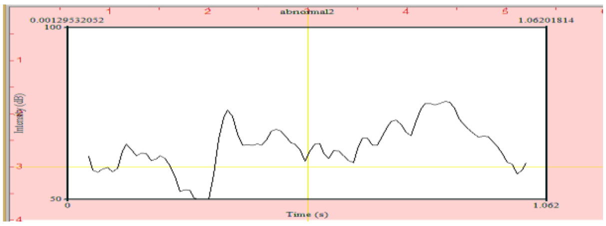


Fig. 29. Intensity spectrum of abnormal voice.

5) Formant frequency analysis

Fig. 26 and Fig. 27 show the formant frequencies of a normal voice and the formant frequencies of an abnormal voice respectively.

6) Intensity analysis

Fig. 28 shows the intensity spectrum of a normal voice which is a regular plot of the voice sample. Fig. 29 shows the intensity spectrum of an abnormal voice which is an irregular plot and does not have a uniform intensity range.

C. Classification of dominant voice parameter

After analysis we can see the most suitable values with which we can classify the given data using KNIME.

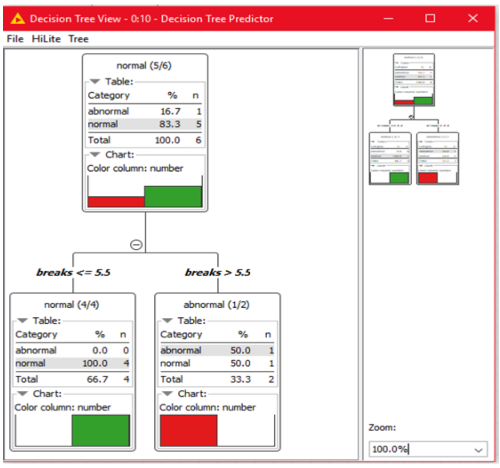


Fig. 30. Parameter classification results for 40%.

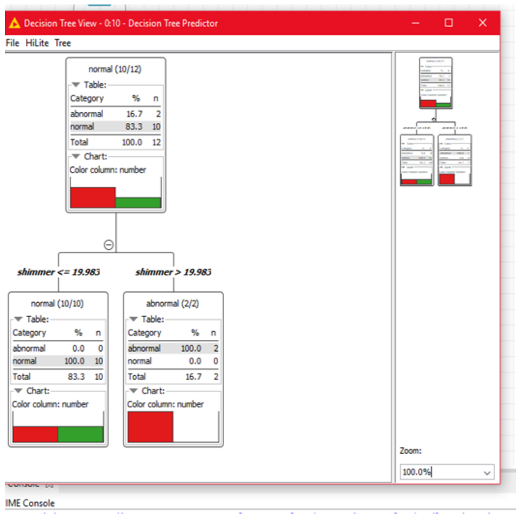


Fig. 31. Parameter classification results for 80%.

1) For relative 40%:

As per the algorithm, the breaks have been used to classify the following data. The more the data the more variables come into play. Here breaks play a role in classifying the data. It is shown in Fig. 30.

2) For relative 80%:

Depending upon the algorithm, the data has been classified using shimmer as the more suitable way possible at the value of 19.983. It is shown in Fig. 31.

Hence after large amount of data and a greater number of iterations the values play a role in classification of the human voice to be normal or abnormal.

IV. CONCLUSION

The voice samples from 50 subjects of male and female equally were analyzed to extract acoustic features of the voice samples using Praat software. The voice feature extraction was tabulated with average values. The dataset is given as the input to the KNIME software to classify the dominant parameter to distinguish the normal and abnormal values of the dataset thereby helping us decide on the patient's abnormality in voice. Although the analysis does not state the individual's voice disorder but will be able to state the degree of severity.

REFERENCES

1. Elizabeth S. Hasseltine, Shannon F. Black, Tayler M. Corcoran, Danika L. DiPalma, Susan E. Dixon, Anne T. Gooch, Lauren M. Hurlburt, Ashton B. Murray, Kathryn B. Potts, Anna C. Schnizler, Caitlin Secrist, Roma Marisa Shickel, Filip Loncke, and Paul Corthals, "Predicting Stuttering Severity Ratings by Timing and Tallying Dysfluencies Using Praat Software," *Contemporary Issues in Communication Science and Disorders*, vol. 43, pp. 106-114, 2016.
2. Mansi Kumbhakarn, and Bageshree Sathe-Pathak, "Analysis of emotional state of a person and its effect on speech features using PRAAT software," *International Conference on Computing Communication Control and Automation*, 2015.
3. Betul Erdogdu Sakar, M. Erdem Isenkul, C. Okan Sakar, Ahmet Sertbas, Fikret Gurgen, Sakir Delil, Hulya Apaydin, and Olcay Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828-834, 2013.

4. Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," *International Conference on Multimodal Interfaces*, pp. 205-211, 2004.
5. Minu George Thoppil, C. Santhosh Kumar, Anand Kumar, and John Amose, "Speech signal analysis and pattern recognition in diagnosis of dysarthria," *Annals of Indian Academy of Neurology*, vol. 20, no. 4, pp. 352-357, 2017.
6. Tan Tze Ern Shannon, Dai Jingwen Annie, and See Swee Lan, "Speech analysis and depression," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016.
7. Haritha C. K., Ana P., Shahanas V. P., and Jenny Mevis D'souza, "Establishment of Acoustic Norms for PRAAT Software in 20-25 Year Old Indian Adults," *Language in India*, vol. 16, no. 8, 2016.
8. Magdalena Majdak, Magdalena Igras, and Anna Domeracka-Kołodziej, "Looking for natural voice-The effectiveness of the program of Postgraduate Studies of Voice and Speech Training," *XXII Annual Pacific Voice Conference (PVC)*, 2014.
9. Bahman Gorjian, Abdolmajid Hayati, and Parisa Pourkhoni, "Using Praat software in teaching prosodic features to EFL learners," *World Conference on Psychology, Counselling and Guidance (WCPCG)*, vol. 84, pp. 34-40, 2013.
10. Lamia Bouafif, and Kais Ouni, "A speech tool software for signal processing applications," *International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, 2012.
11. Sonu, and R. K. Sharma, "Disease Detection Using Analysis of Voice Parameters," *International Journal of Computing Science and Communication Technologies*, vol. 4, no. 2, pp. 416-420, 2012.

AUTHORS PROFILE

Christina Subiksha W received B.Tech degree in Electronics and Communication Engineering from VIT University, Vellore, India in 2019 and currently pursuing MBA in VIT University, Vellore, India. Her areas of interest include signal and image processing.

Nandhini A received B.E degree in Electronics and Communication Engineering from Kingston Engineering College, Vellore, India in 2018 and currently pursuing M.Tech degree in Communication Engineering at VIT University, Vellore, India. Her areas of interest are signal processing and image processing.

Bharath K P received B.E degree in Electronics and Communication Engineering from Amruta Institute of Engineering and Management Sciences, Bangalore, India in 2014 and M.Tech degree in Digital Communication Engineering from Siddaganga Institute of Technology, Tumkur, India in 2016. He is currently pursuing Ph.D. in the School of Electronics Engineering, VIT University, Vellore, India; alongside he is CSIR-Senior Research Fellow, Govt India (From April-2019 to present). He is currently working on speaker identification and speaker verification with various spoofing attacks. His areas of interest are speech processing, signal processing and image processing.

Mahalti Mohammed Sohail currently pursuing B.Tech degree in Electronics and Communication Engineering at VIT University, Vellore, India. His areas of interest are signal processing and digital logic design.

Rajesh Kumar M (SM'16) received B.E degree in Electronics and Communication Engineering from Madras University, Chennai, India in 2000, M.E degree in Electrical Drives and Embedded Control from the College of Engineering, Anna University, Chennai, India in 2005, and M.B.A degree in Systems from Alagappa University, Karaikudi, India in 2008. He was the recipient of the Research Scholarships award for his Ph.D from Northumbria University, Newcastle upon Tyne, U.K. in 2010. He is a Chartered Engineer (with a membership of the IET, U.K.). Since 2000, he has been in teaching with the Department of Electronics and Communication Engineering (India) and Faculty of Engineering and Environment (U.K.). He is currently an Associate Professor with the School of Electronics Engineering, VIT University, Vellore, India. His research interests include image engineering and security, biometrics, speech signal processing, computer vision, perceptual hashing, pattern recognition and optimization technique.