

Screening upper respiratory diseases using acoustics parameter analysis of speaking voice

Abstract

This paper presents the acoustic analysis of speaking voice for screening some of the upper respiratory diseases. The six acoustic parameters of speaking voice were analysed, which were recorded while reading the specific text. The text to read was designed considering phonetic categories of words used and their order of occurrence in the text. The appropriate number of words was combined to form the desired text to read for extracting maximum information from the upper respiratory system. Then recorded speaking voice was analysed for disease-specific corresponding variation in Total Harmonic Distortion (THD), Inter-Modulation Distortion (IMD), Peak Amplitude (PA), Peak Frequency (PF), Signal to Noise Ratio (SNR) and Signal to Noise and Distortion (SIND). This study established the upper respiratory disease-specific signature of speaking voice for Cold & Cough, Pharyngitis, Chronic Suppurative Otitis Media (C.S.O.M.) and Rhinitis.

Keywords: acoustic analysis, human voice, voice analysis, respiratory diseases

Volume 3 Issue 4 - 2017

Santosh Bothe,¹ Monali Bobade²
¹NMIMS University, India

²Daksh Foundation, India

Correspondence: Santosh Bothe, MPSTME, NMIMS University, Shirpur, Dhule, India; Email santoshbothe@gmail.com

Received: September 06, 2017 | **Published:** November 09, 2017

Introduction

Primary disease diagnosing using various biological sound is a practice from ages, where the biological sound called bioacoustics is used for diagnosis. The recent studies have demonstrated that the abnormality due to the disease impacts the voice production system to a significant extent, researchers also studied the wide spectrum of voice pathologies.¹⁻⁵ A significant deviation is observed in speaking voice due to emotions, confidence, feelings and etc.^{6,7} Moreover, analysis of the acoustic parameters of speaking voice can be used for understanding mental and health status.⁸ The focus of the paper was to compare the disease specific variations in acoustic parameter values of the speaking voice and corresponding upper respiratory diseases. Large number studies have been carried out for objective analysis of voice. But most of these studies are focused on analysing the acoustic features which are highly person specific like jitter, pitch, tone using mathematical signal processing techniques.⁹⁻¹¹ There are various frameworks of voice signal analysis being used by researchers like for mood identification, emotion recognition, disease classification etc.,¹²⁻¹⁵ however, many of frameworks that are in practice today, are based on Jitter, Shimmer, Pitch Tone which are highly user specific due to which it cannot be generalised for analysis of individuals.¹⁶

Beside this, voice production system has four main components e.g. motor, vibrator, resonator and articulator. The motor part is responsible for the muscle movement to generate the vibration, vibration are generated by the glottis based on the 'muscle movement trigger' by the nerves system, resonators includes throat, mouth, the nasal and head cavities whereas articulation is performed by the tongue, lips, teeth, hard and soft palate.¹⁷ The normal function of these system depends on the pathology of voice production system as well as the entire body as presence and concentration of different body fluid has a different impact on speaking voice. The script considered for the recording of the voice was designed with an objective that, it will have sufficient words which excite the different motor signal, vibration, resonators and articulation.¹ Another important motivation for the study is the need for the early screening of the disease. As per WHO global health statistics 2014, approximately 10 million people can be saved

each year with an early diagnosis of upper respiratory diseases.¹⁸⁻²⁰ In order to obtain the reliable acoustic signature of "disease-specific variation", voice samples of 931 patients were recorded and analysed of which 502 are male and 429 are female patients and 430 healthy controls of which 229 were male and 201 were female. Our study demonstrated that there are disease specific statistically significant variations in the acoustic parameters of speaking voice. To date, some researchers have more focus on fundamental frequency, pitch contours and jitter, many studies have carried out to study speech rate and intensity differences.²¹⁻²⁴ There are also studies carried out to find segmental features that relate to the precision articulation.^{21,25,26} Scherer and co-researcher studied the impact of speaker's emotion on speech production.²⁷ In a similar research vein, Laukkanen et al.²⁸ identified that anger can be characterised by low quotient values of the glottal flow and sadness, surprise and enthusiasm by having high quotient values.^{28,29} We have tried to map THD, IMD, PF, PA, SNR and SIND of the speaking voice to corresponding variations for upper respiratory diseases.

Voice samples processing system and method

Voice recording and preprocessing

Voice recording was carried out in following position. The subject was in seating position on a comfortable chair, the elbows forming a 90° angle, the arm placed on a front table, in a relaxing condition. The specified text made of words related to the Marathi and English alphabets to obtain the pronunciation of specific vocals and specific consonants. The predefined text giving the, made up by combining Marathi and English words depending on their phonetic signature were used. The underlying idea has been patented by the authors [#0001411389, 20th Oct. 2014, uibm.gov.it].

Selecting the samples

The software pre-processor was developed to eliminate the personal specific characteristics of speaking voice and pre-processed samples were used for further analysis. Diagnosed subjects with upper respiratory diseases were recruited for the study within the ethical

framework of respective hospital. The speaking voice is recorded by “read and repeat” or “listen and repeat” method as per the convenience of the subject. The subjects were recruited from the hospitals in Navi Mumbai region to name, Tata Memorial Hospital (ACTREC, Sector 22, Kharghar, Navi Mumbai - 410208, India.), D. Y. Patil Hospital (Sector 5, Nerul Navi Mumbai, Maharashtra 400614).

Hardware

Uni-directional Logitech microphone with characteristics: 8-28000Hz, -59dBV/μBar, -39dBV/Pa+/-4dB is used for recording the voice samples. The microphone was plugged into a standard personal computer running windows XP.

Software

The “Audalysis” software was coded in visual basic 6 to extract the voice parameter of interest from the recorded voice. The recording performed at a sample rate of 41000Hz and saved in .WAV format. Audalysis is modular, user-friendly and does not require any special training to use it. The software allows selecting the sample rate (6000Hz, 8000Hz, 11025Hz, 22050Hz, or 44100Hz), the number of the audio channels (1 or 2) and the resolution of bits (8 or 16). For the current work, we selected 44100Hz, 2 channels and 16bits. These functions are shown in the block diagram (Figure 1).

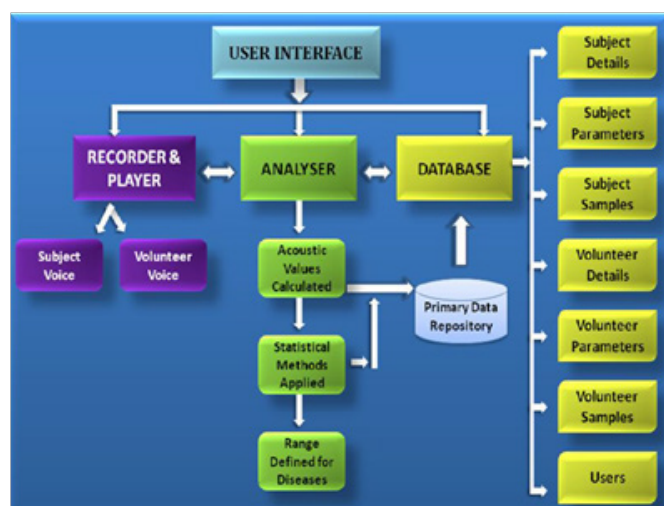


Figure 1 Audalysis is the home-made software used to record, store and process the signal voice parameters. Its modular architecture consists of three main blocks: the recorder/player, the analyser and the database.

The equations of the parameter used

Total harmonic distortion (THD)

THD is the ratio of the sum of the powers of all harmonic frequencies above the fundamental frequency to the power of the fundamental frequency. It compares the output signal of the amplifier with the input signal and measures the level differences in harmonic frequencies between the two. It is computed by searching the entire spectrum to find the peak frequency (fundamental) and then calculating the total power in the harmonic frequencies. The THD level is then computed as the ratio of the total harmonic power to the fundamental power. Residual noise is not included in this calculation Formula:

$$THD = \frac{\sum \text{harmonic powers}}{\text{Fundamental frequency powers}} = \frac{P_2 + P_3 + \dots + P_n}{P_1} \quad \dots \dots \text{(Eqn 1)}$$

Equation 1: Calculation of Total Harmonic Distortion

Peak frequency

The peak frequency utility will display the frequency of the strongest spectral component in the entire span. “Higher pitched” or “lower pitched” sounds are pressure vibrations having a higher or lower number of cycles per second. Formula is given by:

$$f / \frac{1}{T} \quad \dots \dots \text{(Eqn 2)}$$

Equation 2: Calculation of Peak Frequency.

Where T is the Period, which is reciprocal of frequency.

Peak amplitude

The peak amplitude is the amplitude of the strongest spectral component in the entire span. Amplitude is the magnitude of change in the oscillating variable, with each oscillation, within an oscillating system. If the variable undergoes regular oscillations, and a graph of the system is drawn with the oscillating variable as the vertical axis and time on the horizontal axis, the amplitude is visually represented by the vertical distance between the extremes of the curve. Peak amplitude is measured between a peak and a rest position of the system.

It is given by the formula:

$$X = A \sin(t - K) + b \quad \dots \dots \text{(Eqn 3)}$$

Equation 3: Calculation of peak amplitude

Where, A is the amplitude, t: time, K and b are arbitrary constants for time and displacement offsets respectively.

Signal to noise and distortion (SINAD)

SINAD is a common sensitivity measurement, it is an acronym for signal plus noise and distortion and is equal to (S+N)/N in our system. It is expressed in dB. If the signal is much stronger than the noise then SINAD value approaches the SNR value. Otherwise, SINAD will be greater than the SNR. The ratio is expressed as a logarithmic value (in dB) from the formulae $10 \log(SND/ND)$. It can be summarized as the ratio of the total signal power level (Signal + Noise + Distortion) to unwanted signal power (Noise + Distortion). Accordingly, the higher the figure for SINAD, the better the quality of the audio signal. The SINAD figure is expressed in decibels (dB) and can be determined from the simple formula:

$$SINAD = 10 \log \left(\frac{SND}{ND} \right) \quad \dots \dots \text{(Eqn 4)}$$

Equation 4: Signal to Noise and Distortion.

Where: SND = combined Signal + Noise + Distortion power level.

ND = combined Noise + Distortion power level.

Inter-modulation distortion (IMD)

Inter-Modulation Distortion (IMD) is a measure of the distortion caused by the interaction (mixing) of two tones. When multiple signals are injected into a voice, may be due to the presence of multiple diseases undesired modulations or mixing of these two signals can occur. The IMD level is calculated by first computing the frequencies and amplitudes of the two strongest tones in the spectrum which helps in finding the association even in a case of co-existence of the disease-specific variations in pathologies. The total power in each of the intermodulation product frequencies is then computed. IMD is the ratio of the intermodulation power to the RMS sum of the tone power.

Signal to noise ratio (SNR)

The Signal to Noise Ratio (SNR) is the ratio of the signal peak power level to the total noise level and is expressed in decibels (dB). It represents the maximum capacity of various organs specific to the text being pronounced. The SNR is computed by searching the entire spectrum to find the peak frequency and then calculating the total noise power in the remaining spectrum. The SNR is then computed as the ratio of the noise power to the peak power and expressed in decibels.

It is given by the formula:

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) = 20 \log_{10} \left(\frac{A_{\text{signal}}}{A_{\text{noise}}} \right) \quad \dots (\text{Eqn 5})$$

Equation 5: Signal to Noise Ratio

Where, P is average power and A is RMS amplitude.

Sampling rate.

Recording system

The Sampling Rate determines how many times a second the analogue input signal is “sampled” or digitized by the sound card. In our study sample rate of 41000 Hz was used considering the highest frequencies of audible range 20000Hz and “Nyquist Sampling Theorem” which states that any signal can be represented if sampled at least twice the rate of the highest frequency of interest.

Recording screen

There are specific scripts based on the phonation association with the specific organs. The primary clinical diagnosis was used for the selecting the appropriate script and the orders of the words in the script. The speaking voice was recorded using the Audalysis software developed by the authors in Visual Basic 6 (Figure 2).

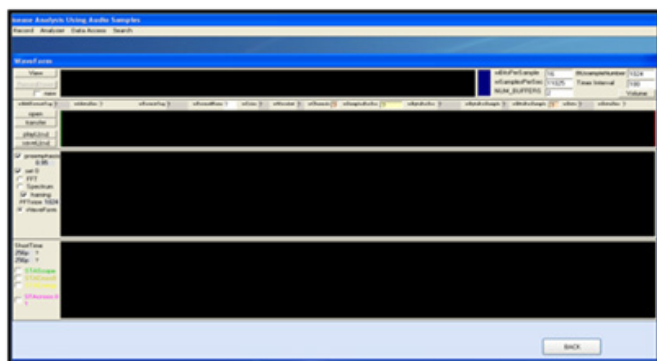


Figure 2 Snapshot of the recording screen of “Audalysis”.

Recording time duration

The time duration of 45 seconds has been considered for a recording of the voice samples sufficient for exciting the relevant systems. This time, duration also serves as optimum for a recording of peaks in the human voice.³⁰ Hence, voice recorded in 45 seconds helps in getting optimum results for the considered parameters. The rule followed while recording was, do not stop in between the script reading, and allow to complete the script even if it more than 45 seconds, if it is less than 45-second repeat entire script to cross 45 seconds, generally two repetitions were common to reach 45 seconds. Every patient was requested to rehearse the script so that they are able to read without

fumbling. Each patient was relaxed, seated on a comfortable chair and the microphone 6-8cm apart from the mouth.

Results and discussion

The Table 1 below shows the number of diagnosed patients for the different upper respiratory disease recruited for the study. In statistical analysis, following assumptions were taken,

- The sampling distribution is normally distributed.
- Samples were collected with random samples.
- According to the data, $\pm 5\%$ flexibility was introduced to calculate the mean range.

Statistical analysis in Table 1 clearly depict that significant difference is observed in Peak Frequency, IMD and THD parameters for upper respiratory disorders, a significant difference was observed in the mean values of PF, SINAD, IMD and THD whereas PA and SNR shows minor deviation, with respect to the standard reference range obtained for healthy controls. A major difference was seen in PF and IMD, it is less significant to discriminate between cold & Cough and Pharyngitis. The difference for CSOM and Rhinitis was also significant. The results were compared with the predefined reference range of parameter established by recording the sample of the healthy controls¹ (Table 2 & Table 3). The mean range was calculated as a measure to describe the central tendency of the data. The sample mean makes a good estimator of the population mean as its expected value is the same as the population mean. Mean range is used to eliminates errors. There was a significant difference in the mean of the acoustic parameters between the subject and the healthy volunteer.

Table 1 Number of patients recruited for the study of upper respiratory disease

Disease	Cold & cough		Pharyngitis		C.S.O.M		Rhinitis	
Total	45		67		23		56	
Gender	M	F	M	F	M	F	M	F
Gender total	21	24	32	35	16	7	32	24

Table 2 Arithmetic mean (upper respiratory disorder) where column 1: Acoustic Parameters, 2: Healthy Control, 3: Cold and & Cough, 4: Pharyngitis, 5: CSOM and 6: Rhinitis

I *	2*	3*	4*	5*	6*
PF (Hz)	450.2041	245.0119	247.4353	371.3298	314.7561
PA (dB)	-42.8954	-42.0246	-40.5069	-40.5088	-40.3763
SINAD	1.9189	1.5928	1.3895	1.5591	1.4621
IMD	538.9156	71.0669	72.2578	60.1324	62.9172
SNR	-3.3673	-4.0102	-4.5781	-4.0942	-4.3828
THD	128.5062	102.878	97.8426	78.4088	79.5227

Interpretation

Though there are overlapping mean range for the Cold & Cough and Pharyngitis, the individual signature of the disease in 3-dimensional analysis with respect to disease, variation pattern and acoustic parameters shows unique behaviour corresponding to each disease considered for the study. In the Post Hoc results means for groups in homogeneous subsets are calculated and the probability

values of all the acoustic parameters segregated on the basis of diseases using harmonic mean sample size=1077.835, following observations was made. Healthy control data showed a marked difference in the probability values for Peak Frequency as compared to all the subject data, upper respiratory disorders show a significant probability range for PF, SINAD, IMD and THD distinguishing it from other diseases. We have evaluated the relationship between disease and a standardized reference value (obtain by analysing healthy control) comparing acoustic measurements made during the diseased state and the normal state of the different subjects. The voice data collected from all subjects and healthy control was compared and statistical methods (CI

and T-Test) were used for further evaluation, The t-test analysis results above in the table confirms significant difference (since significance values are less than that of 0.05) observed in the arithmetic means of PF, PA, SINAD and IMD parameters, Which means that PF, PA, SINAD and IMD values can be used for predicting upper respiratory disease up to a certain extent. Also taking a threshold as 0.05, values which are significantly below this range have a significant probabilistic variation and are therefore considered. Values above this threshold were rejected. Upper respiratory tract except SNR and THD, values are above 0.05 and thus, the values are accepted.

Table 3 Arithmetic mean range for upper respiratory diseases for PF, PA, SIND IMD, SNR and THD. Where column 1: Acoustic Parameters, 2: Healthy Control, 3: Cold and Cough, 4: Pharyngitis, 5: CSOM and 6: Rhinitis

1	2	3	4	5	6
PF (Hz)	427.693 to 472.714	241.459 to 248.565	352.763 to 389.896	299.018 to 330.493	235.063 to 259.807
PA (dB)	-40.751 to -45.041	-39.923 to -44.126	-38.483 to -42.534	-38.357 to -42.395	-38.482 to -42.532
SINAD	1.9093 to 1.9285	1.5132 to 1.6724	1.4811 to 1.6371	1.389 to 1.535	1.3201 to 1.4589
IMD	511.971 to 565.861	67.5136 to 74.620	57.125 to 63.139	59.771 to 66.063	68.644 to 75.871
SNR	-3.1991 to -3.5361	-3.8096 to -4.2107	-3.889 to -4.298	-4.1636 to -4.6019	-4.3492 to 4.80701
THD	122.081 to 134.932	97.734 to 108.022	74.488 to 82.329	75.546 to 83.498	92.950 to 102.735

Conclusion

It is clear that voice sample analytics may be applied to as an aid to clinical screening and diagnosis. The techniques are handy and portable. It possible to perform useful investigations using standard personal computers, as installed in most of the clinicians consulting room, and can, therefore, be of particular value to primary care physicians who do not have easy access to sophisticated diagnostic equipment. Though the project is in an early stage, however, it has a potential of being reliable clinical screening technology after further research. The most likely area for early exploitations is generating the personalised voice signature database to have a person specific signature and remote monitoring of diseases. An exciting prospect for the future would the routine availability of a miniaturized portable apparatus with the ability to capture both sound and airflow, implement simple and clinically useful analysis packages and when necessary, communicate data via mobile telephony to a specialist centre in a local hospital. The important factor of voice production system e.g. motor movement of vocal muscles, glottis vibration, respective nerves system trigger for various motor movements, absorption coefficient and resonance are dependent on the pathology of a human body as a whole. The fundamental frequency produced by glottis is dependent on the biochemical composition of the glottis, the presence and absence of certain body fluids impact the vibration capacity of the glottis e.g. fundamental frequency, resonance and absorption coefficients of resonators and articulation system, thereby it impacts acoustic characteristics of speaking voice. Still further clinical and pathological validations are required to validate the obtained disease specific acoustic signatures.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Santosh Bothe, Giovanni Saggio. Relevance of voice analysis in diagnosing tuberculosis. Conference WVITAE; India: 2015.
2. Gobl N Ailbhe. The role of voice quality in communicating emotion, mood and attitude. *Speech communication*. 203;40(1-2):189–212.
3. Scherer KR. Vocal communication of emotion: A review of research paradigms. *Speech communication*. 2003;40(1-2):227–256.
4. Ververidis D, Kotropoulos C. Emotional speech recognition: Resources, features, and methods. *Speech communication*. 2006;48(9):1162–1181.
5. Arias-Londoño JD, Godino-Llorente JJ, Sáenz-Lechón N, et al. An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern recognition*. 2010;43(9):3100–3112.
6. Arjmandi MK, Pooyan M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomedical Signal Processing and Control*. 2012;7(1):3–19.
7. Godino-Llorente JJ, Sáenz-Lechón N, Osma-Ruiz V, et al. An integrated tool for the diagnosis of voice disorders. *Medical engineering & physics*. 2006;28(3):276–289.
8. Gloria S Watersa, Elizabeth Rochon, David Caplan. Task demands and sentence comprehension in patients with dementia of the Alzheimer's. *Journal of brain and language*. 1998;62(3):361–397.
9. Niedzielska G. Acoustic analysis in the diagnosis of voice disorders in children. *Int J Pediatr Otorhinolaryngol*. 2001;57(3):189–193.
10. Alpan J, Schoentgen J, Maryn Y, et al. Assessment of disordered voice via the first harmonic. *Speech communication*. 2012;54(5):655–663.
11. Teija Waaramaa, Anne-Maria Laukkanen, Matti Airas, et al. Perception of emotional valences and activity levels from vowel segments of continuous speech. *Journal of voice*. 2010;24(1):30–38.
12. Smiljanić R, Bradlow AR. Temporal organization of English clear and conversational speech. *J Acoust Soc Am*. 2008;124(5):3171–3182.

13. Klaus R Scherer, Rainer Banse, Harald G Wallbott. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of cross-cultural psychology*. 2001;32(1):76–92.
14. Tsanas A, Little MA, Mc Sharpy PE, et al. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng*. 2012;59(5):1264–1271.
15. Kappiarukudil KJ, Ramesh MV. Real-time monitoring and detection of Heart Attack using wireless sensor networks. *IEEE Proceedings of the fourth international conference on Sensor technologies and applications at Venice, Italy: Springer*; 2010. p. 632–636.
16. Thomas Fillebrown. Resonance in Singing and Speaking. 2006. p. 7–32
17. Roark RM. Frequency and voice: Perspectives in the time domain. *Journal of Voice*. 2006;20(3):325–354.
18. Tsanas A, Little MA, Mc Sharpy PE, et al. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc Interface*. 2011;8(59):842–855.
19. Gonzales R, Steiner JF, Sande MA. Antibiotic Prescribing for Adults With Colds, Upper Respiratory Tract Infections, and Bronchitis by Ambulatory Care Physicians. *JAMA*. 1997;278(11):901–904.
20. Yadollahi ZM, Moussavi. Acoustical respiratory flow. *Engineering in Medicine and Biology Magazine*. 2007;26(1):56–61.
21. Scherer KR, Darby J. Speech and emotional states. *The Evaluation of Speech in Psychiatry and Medicine*. USA: Grune and Stratton; 1981. p. 189–220.
22. Scherer KR. Vocal affect expression: A review and a model for future research. *Psychol Bull*. 1986;99(2):143–165.
23. Scherer KR. *Vocal measurement of emotion*. In: Plutchik R et al. editors. *Emotion: Theory, Research, and Experience*, USA: Academic Press; 1986;4:233–259.
24. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *J Acoust Soc Am*. 1972;52(4):1238–1250.
25. Strik H, Boves L. On the relation between voice source parameters and prosodic features in connected speech. *Speech Communication*. 1992;11(2-3):167–174.
26. Scherer KR. *Affect bursts*. In: Van Goozen SHM, et al. editors. *Emotions*, Lawrence Erlbaum, USA: Springer; 1994. p. 161–193.
27. Laukkanen AM, Vilkmán E, Alku P, et al. *On the perception of emotional content in speech*. Sweden: Proceedings of the XIIIth International Congress of Phonetic Sciences; 1995. p. 246–249.
28. Laukkanen AM, Vilkmán E, Alku P, et al. Physical variation related to stress and emotionally state: a preliminary study. *Journal of Phonetics*. 1996;24(3):313–335.
29. Tsanas A, Little MA, Mc Sharpy PE, et al. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE Transactions on Biomedical Engineering*. 2009;57(4):884–888.
30. Uloza V, Verikas A, Bacauskiene M, et al. Categorizing normal and pathological voices: automated and perceptual categorization. *Journal of Voice*. 2011;25(6):700–708.