THE USE OF ARTIFICIAL INTELLIGENCE IN DIAGNOSING BENIGN AND MALIGNANT LARYNGEAL DISORDERS

Dr. Neha Chaudhary

Junior Resident

Dept. of ENT, GSVM Medical College, Kanpur

ABSTRACT

Introduction: Voice disturbances are among the earliest manifestations of laryngeal pathologies. While benign lesions typically cause mild dysphonia, malignant disorders often produce profound and progressive alterations in vocal quality. Differentiating benign from malignant laryngeal disease at an early stage is critical for timely diagnosis and management. Acoustic analysis offers objective, non-invasive, and reproducible evaluation of voice. This study aimed to compare acoustic parameters between benign and malignant laryngeal disorders and to identify robust markers capable and implementing them in an AI based software for diagnosing the disease without any invasive procedures.

Materials and Methods: This retrospective observational study included 38 patients: 19 with histopathologically confirmed benign laryngeal lesions and 19 with laryngeal carcinoma. Acoustic analysis was performed on sustained phonation samples using standardized software. Parameters studied included fundamental frequency (F0 mean, median, min, max, range), perturbation measures (jitter local, jitter RAP, shimmer local, shimmer APQ3, shimmer APQ5), cepstral peak prominence (CPP), harmonic-to-noise ratio (HNR), intensity range, spectral measures (centroid, flatness, zero-crossing rate), and long-term average spectrum (LTAS) in multiple frequency bands. Statistical analysis involved Shapiro–Wilk testing for normality, Welch's t-test or Mann–Whitney U test as appropriate, and Bonferroni correction for multiple comparisons across 26 variables. A p-value < 0.05 was considered significant.

Results: Malignant voices exhibited greater perturbation, reduced harmonic stability, and spectral degradation compared to benign voices. Several parameters showed raw significance ($p < 0.05$), including F0 median, F0 range, jitter RAP, shimmer local, shimmer APQ3, shimmer APQ5, CPP, and LTAS (0–1 kHz). After Bonferroni correction, five parameters remained statistically significant: F0 Max ($p=0.00018$), Shimmer APQ3 ($p=0.0022$), Shimmer APQ5 ($p=0.0016$), CPP ($p=3.85\times10^{-6}$), and LTAS 0–1 kHz ($p=0.0030$).

Conclusion: This study identifies F0 Max, Shimmer APQ3, Shimmer APQ5, CPP, and LTAS 0–1 kHz as robust acoustic markers differentiating malignant from benign laryngeal disorders. These parameters hold potential as objective, non-invasive adjuncts for early diagnosis and monitoring of malignant voice pathology, complementing conventional clinical evaluation.

Keywords: Laryngeal cancer, benign lesions, acoustic analysis, shimmer, cepstral peak prominence, LTAS, Artificial Intelligence, AI based cancer diagnosis, AI in cancer.

INTRODUCTION

Voice is a critical component of human communication, and its alteration often provides the earliest clinical clue in laryngeal pathologies. Benign lesions such as nodules, polyps, and cysts typically cause mild to moderate dysphonia, while malignant laryngeal disorders frequently result in more profound and progressive voice changes. Early differentiation between benign and malignant conditions is essential for timely diagnosis and management, particularly given the prognostic importance of early-stage laryngeal cancer detection [1,2].

Acoustic analysis provides objective and reproducible markers of voice quality. Parameters such as fundamental frequency (F0), jitter, shimmer, harmonic-to-noise ratio (HNR), cepstral peak prominence (CPP), and long-term average spectrum (LTAS) have been widely studied [3–6]. Previous studies suggest malignant voice disorders are associated with greater perturbation, spectral distortion, and reduced harmonic stability compared to benign disorders [7–9]. However, findings have been inconsistent across populations and methodologies.

This study systematically compared acoustic voice parameters between benign and malignant laryngeal disorders, applying rigorous statistical methods with correction for multiple comparisons to identify robust discriminators.

STUDY OBJECTIVES

1. To compare acoustic parameters (F0-related, perturbation measures, spectral features, LTAS) between benign and malignant laryngeal disorders.
2. To determine which variables show statistically significant differences after correction for multiple testing.
3. To evaluate the diagnostic potential of acoustic markers in clinical voice evaluation.

MATERIALS & METHODS

Study Design and Participants

This was a retrospective observational study conducted on voice recordings of patients with histopathologically confirmed laryngeal pathologies. Two datasets representing benign (n=19) and malignant (n=19) cases were included.

Acoustic Analysis

Voice samples were analyzed with standardized software (PRAAT). Parameters examined:

- Frequency/amplitude: RMS, dB, F0 mean, F0 median, F0 min, F0 max, F0 range.
- Perturbation: Jitter Local, Jitter RAP, Shimmer Local, Shimmer APQ3, Shimmer APQ5.
- Voice quality: HNR, CPP, Intensity range.
- Spectral: Spectral centroid, spectral flatness, Zero Crossing Rate (ZCR).
- LTAS: 0–1 kHz, 1–2 kHz, 2–4 kHz, 4–8 kHz.

Statistical Analysis:

- Shapiro–Wilk test for normality.
- Welch t-test for normally distributed data.
- Mann–Whitney U test for non-parametric distributions.
- Bonferroni correction for multiple comparisons across 26 variables.
- Significance threshold: $p < 0.05$.

Legend: Acoustic Voice Parameters

Basic Parameters

- Filename → Name of the audio file (patient identifier).
- Duration (s) → Length of the analyzed segment in seconds (20s maximum used here).
- RMS → Root Mean Square amplitude; reflects overall loudness.
- dB → Decibel level, converted from RMS amplitude [$dB = 20 \cdot \log_{10}(RMS)$].

Pitch-Related

- F0 Mean / Median / Min / Max (Hz) → Fundamental frequency (perceived as pitch). Mean, median, lowest and highest pitch values across the sample.
- F0 SD (Hz) → Standard deviation of pitch; measures pitch variability.
- F0 Range (Hz) → Highest minus lowest pitch; shows vocal flexibility.

Perturbation Measures (Cycle-to-Cycle Instability)

- Jitter Local → Variation in fundamental period between cycles; reflects pitch instability.
- Jitter RAP (Relative Average Perturbation) → Average of short-term pitch variations over 3 cycles (smoother jitter measure).
- Shimmer Local → Variation in amplitude between cycles; reflects loudness instability.
- Shimmer APQ3 / APQ5 → Amplitude Perturbation Quotients over 3 or 5 cycles; smoother shimmer measures.

Intensity

- Intensity Range (dB) → Difference between loudest and softest points in the sample.

Voice Quality / Spectral Features

- HNR Approx (dB) → Harmonic-to-Noise Ratio; higher values = clearer, less breathy voice.
- Spectral Centroid (Hz) → “Center of gravity” of spectrum; higher = brighter or sharper voice quality.
- Spectral Flatness (0–1) → How noise-like the spectrum is (0 = tonal, 1 = noisy).

- Spectral Roll-off 85% (Hz) → Frequency below which 85% of energy is contained.

Formant Frequencies (Resonances)

- F1, F2, F3 (Hz) → The first three formant frequencies (resonance peaks of the vocal tract); important for vowel quality.

Cepstral / Nonlinear Measures

- CPP (Cepstral Peak Prominence) → Measures voice quality/dysphonia; higher = clearer, more periodic voice.
- ZCR (Zero-Crossing Rate) → Rate at which the signal waveform crosses zero; higher values suggest more noise.

Long-Term Average Spectrum (LTAS)

- LTAS 0–1k %, 1–2k %, 2–4k %, 4–8k % → Percentage of spectral energy in these frequency bands; indicates how energy is distributed across low vs high frequencies.

RESULTS

Variable analysis:

1. RMS:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.1010	0.0255	0.1053	0.0211	0.0318-0.1383
Malignant	19	0.0922	0.0440	0.0833	0.0569	0.0300-0.1930

Normality (Shapiro-Wilk): Benign p=0.056, Malignant p=0.468, Test: Welch t-test (t=0.752)

Raw p=0.458 → Bonferroni p=1.000

Significant: No

2. dB (sound level dB):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	-20.29	2.92	-19.60	1.80	-29.90 - -17.18
Malignant	19	-21.74	4.40	-21.58	6.25	-30.40 - -14.30

Normality: Benign p=0.0003, Malignant p=0.914, Test: Mann–Whitney U (U=215)

Raw p=0.321 → Bonferroni p=1.000

Significant: No

3. Intensity range (dB):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	17.66	2.05	18.10	3.75	14.3-20.3
Malignant	19	18.28	2.86	18.70	3.37	12.9-22.3

Normality: Benign p=0.031, Malignant p=0.165, Test: Mann–Whitney U (U=154)

Raw p=0.448 → Bonferroni p=1.000

Significant: No

4. F0 mean (Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	214.67	31.54	217.10	37.15	146.4-282.7
Malignant	19	200.51	37.91	202.76	37.18	114.8-268.2

Normality: Benign p=0.992, Malignant p=0.646, Test: Welch t-test (t=1.252)

Raw p=0.219 → Bonferroni p=1.000

Significant: No

5. F0 median (Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	198.56	43.09	200.50	60.20	124.6-290.1
Malignant	19	167.30	41.52	163.64	36.70	100.0-263.6

Normality: Benign p=0.982, Malignant p=0.294, Test: Welch t-test (t=2.277)

Raw p=0.0288 → Bonferroni p=0.750

Significant: Raw = Yes, Bonferroni = No

6. F0 min (Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	68.05	7.17	66.40	5.25	60.6-93.4
Malignant	19	76.10	12.65	74.07	19.52	60.6-104.2

Normality: Benign p=0.00016, Malignant p=0.163, Test: Mann–Whitney U (U=118)

Raw p=0.0702 → Bonferroni p=1.000

Significant: No

7. F0 max (Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	501.10	~0	501.10	0	501.1-501.1
Malignant	19	441.29	100.10	500.00	91.65	235.3-501.1

Normality: Benign p=1.000, Malignant p=8.66e-06, Test: Mann–Whitney U (U=323)

Raw p=6.99e-06 → Bonferroni p=0.00018

Significant: Raw = Yes, Bonferroni = Yes

8. F0 range:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	433.09	7.19	434.70	5.25	407.7-440.6
Malignant	19	365.18	101.42	417.36	111.71	157.8-439.4

Normality: Benign p=0.00016, Malignant p=8.39e-05, Test: Mann–Whitney U (U=274)

Raw p=0.00661 → Bonferroni p=0.172

Significant: Raw = Yes, Bonferroni = No

9. Jitter local:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.2482	0.0513	0.2390	0.068	0.173-0.371
Malignant	19	0.2719	0.1470	0.2579	0.244	0.067-0.499

Normality: Benign p=0.339, Malignant p=0.125, Test: Welch t-test (t=-0.666)

Raw p=0.512 → Bonferroni p=1.000

Significant: No

10. Jitter RAP:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.0107	0.0028	0.0106	0.0035	0.0068-0.0193
Malignant	19	0.0134	0.0091	0.0111	0.0091	0.0065-0.0547

Normality: Benign p=0.262, Malignant p=0.031, Test: Mann–Whitney U (U=129)

Raw p=0.0382 → Bonferroni p=0.993

Significant: Raw = Yes, Bonferroni = No

11. Shimmer local:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.0356	0.0115	0.0333	0.0153	0.0126-0.0634
Malignant	19	0.0470	0.0226	0.0428	0.0257	0.0195-0.1228

Normality: Benign p=0.012, Malignant p=0.00016, Test: Mann–Whitney U (U=87)

Raw p=0.00357 → Bonferroni p=0.0925

Significant: Raw = Yes, Bonferroni = No

12. Shimmer APQ3:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.0176	0.0062	0.0160	0.0072	0.0063-0.0328
Malignant	19	0.0266	0.0126	0.0251	0.0146	0.0103-0.0597

Normality: Benign p=0.090, Malignant p=0.00012, Test: Mann–Whitney U (U=64)

Raw p=8.63×10^{-5} → Bonferroni p=0.00224

Significant: Raw = Yes, Bonferroni = Yes

13. Shimmer APQ5:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.0228	0.0076	0.0211	0.0106	0.0078-0.0448
Malignant	19	0.0340	0.0137	0.0331	0.0177	0.0130-0.0739

Normality: Benign p=0.161, Malignant p=6.09×10^{-5}, Test: Mann–Whitney U (U=61)

Raw p=5.95×10^{-5} → Bonferroni p=0.00155

Significant: Raw = Yes, Bonferroni = Yes

14. Spectral Centroid (Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	2441.79	664.62	2419.10	929.95	1012.1-4200.3
Malignant	19	2296.84	730.55	2046.80	1024.65	960.7-4228.4

Normality: Benign p=0.283, Malignant p=0.189, Test: Mann–Whitney U (U=154)

Raw p=0.448 → Bonferroni p=1.000

Significant:No

15. Spectral Flatness:

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.0789	0.0290	0.0722	0.0299	0.0477-0.1535
Malignant	19	0.1057	0.0587	0.0920	0.0753	0.0497-0.3086

Normality: Benign p=0.00017, Malignant p=0.00363, Test: Mann–Whitney U (U=113)

Raw p=0.0147 → Bonferroni p=0.383

Significant: Raw = Yes, Bonferroni = No

16. CPP (Cepstral Peak Prominence):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	15.86	2.82	16.10	3.25	10.8-20.6
Malignant	19	11.53	3.20	11.20	4.25	6.2-18.2

Normality: Benign p=0.000145, Malignant p=0.000268, Test: Mann–Whitney U (U=23)

Raw $p=1.48\times10^{-7}$ → Bonferroni $p=3.85\times10^{-6}$

Significant: Raw = Yes, Bonferroni = Yes

17. ZCR (Zero Crossing Rate):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	0.0501	0.0148	0.0474	0.0185	0.0229-0.0891
Malignant	19	0.0517	0.0221	0.0462	0.0253	0.0173-0.1084

Normality: Benign p=0.0178, Malignant p=0.0258, Test: Mann–Whitney U (U=166)

Raw p=0.570 → Bonferroni p=1.000

Significant: No

18. LTAS (0-1k Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	46.88	17.01	44.70	19.40	13.2-82.6
Malignant	19	31.63	14.85	30.70	20.05	10.7-63.5

Normality: Benign p=0.00344, Malignant p=0.000159, Test: Mann–Whitney U (U=42)

Raw $p=1.16\times10^{-4}$ → Bonferroni p=0.00303

Significant: Raw = Yes, Bonferroni = Yes

19. LTAS (2-4k Hz):

Group	n	Mean	SD	Median	IQR	Min-Max
Benign	19	12.81	5.85	11.30	7.85	5.2-29.6
Malignant	19	20.03	8.04	19.20	9.45	7.7-39.1

Normality: Benign p=0.000165, Malignant p=8.96×10, Test: Mann–Whitney U (U=135)

Raw p=0.124 → Bonferroni p=1.000

Significant: No

Malignant voices consistently demonstrated greater perturbation, spectral distortion, and reduced harmonic energy.

- **Clinically significant (RAW omly):** ($p < 0.05$): F0 Median, F0 Max, F0 Range, Jitter RAP, Shimmer Local, Shimmer APQ3, Shimmer APQ5, Spectral Flatness, CPP, LTAS (0–1 kHz).
- **Bonferroni significance (adjusted $p < 0.05$):**
    - F0 Max (Hz) (adj. p=0.00018)
    - Shimmer APQ3 (adj. p=0.0022)
    - Shimmer APQ5 (adj. p=0.0016)
    - CPP (adj. $p=3.85\times10^{-6}$)
    - LTAS (0–1 kHz) (adj. p=0.0030)

These five emerged as robust discriminators between benign and malignant groups.

DISCUSSION

This study demonstrates that malignant laryngeal disorders exhibit measurable deterioration across acoustic parameters.

- F0 Max reduction reflects limited vibratory range and stiffness due to tumor infiltration [10].
- Shimmer APQ3/5 elevation indicates greater amplitude irregularity, consistent with pathological vibratory asymmetry [11].
- CPP depression is strongly linked to dysphonia severity and voice quality impairment [12,13].
- LTAS (0–1 kHz) reduction suggests diminished low-frequency harmonic energy, reflecting breathy, unstable phonation in malignancy [14].

Interestingly, several other parameters reached raw significance though they did not remain significant after Bonferroni correction. These included F0 Median, F0 Range, Jitter RAP, Shimmer Local, and Spectral Flatness. While these findings must be interpreted with caution due to the risk of type I error, their consistent trend towards worsening in malignant voices suggests they may still hold clinical relevance, particularly in exploratory or adjunctive settings. In larger datasets with more statistical power, some of these measures might achieve robust significance.

Thus, shimmer (especially APQ subtypes), CPP, and LTAS emerge as strong markers, while perturbation and spectral measures may contribute to broader multi-parametric diagnostic models.

Although jitter measures and spectral flatness showed raw differences, they did not survive correction, highlighting the need for stringent statistical validation.

Clinical Implications

- CPP and shimmer measures may serve as reliable, non-invasive biomarkers for early malignancy detection.
- Acoustic analysis can complement laryngoscopy in screening and monitoring.

Limitations

- Some parameters (HNR) lacked variability.
- Cross-sectional design limits longitudinal insights.

Future studies should validate these markers in larger, prospective cohorts and integrate them with machine learning approaches for diagnostic support.

SUMMARY & CONCLUSION

Among 26 acoustic parameters analyzed, five (F0 Max, Shimmer APQ3, Shimmer APQ5, CPP, LTAS 0–1 kHz) demonstrated statistically significant differences between benign and malignant laryngeal disorders after multiple comparison correction.

Additionally, F0 Median, F0 Range, Jitter RAP, Shimmer Local, and Spectral Flatness showed raw statistical significance, suggesting potential as supportive markers. While not definitive after stringent adjustment, these parameters may still contribute meaningfully when combined into composite acoustic indices or machine learning models.

These findings highlight the diagnostic potential of acoustic analysis as an objective adjunct to clinical voice evaluation, particularly in early identification of malignant laryngeal disease.

This study evaluated 26 acoustic parameters to differentiate between benign and malignant laryngeal disorders. After rigorous statistical analysis with Bonferroni correction, five parameters (F0 Max, Shimmer APQ3, Shimmer APQ5, CPP, and LTAS 0–1 kHz) were identified as robust and statistically significant discriminators. These variables consistently reflected the physiological impact of malignant disease, such as reduced vibratory capacity, increased irregularity of vocal fold oscillation, diminished harmonic energy, and compromised spectral structure. The persistence of significance after correction underscores their reliability as diagnostic markers.

In addition to these robust markers, several other parameters — F0 Median, F0 Range, Jitter RAP, Shimmer Local, and Spectral Flatness — demonstrated raw statistical significance. While these did not withstand stringent correction, they remain noteworthy because they highlight consistent trends toward worsening in malignant voices. These findings suggest that although such parameters may lack robustness individually, they could still contribute diagnostic value in larger datasets or when integrated into multi-parametric models. This distinction between robust and exploratory markers emphasizes the importance of cautious interpretation but also encourages continued investigation into broader acoustic profiles of malignancy.

From a clinical standpoint, the identification of objective, reproducible acoustic markers holds considerable potential. These findings reinforce the role of acoustic analysis as a valuable adjunct to conventional laryngoscopy and stroboscopy. In routine practice, voice analysis could serve as an accessible screening tool for high-risk populations, allowing earlier identification of malignant disease and more timely interventions. Moreover, its non-invasive nature makes it particularly suitable for repeated assessments during follow-up, facilitating longitudinal monitoring of disease progression and treatment outcomes.

Beyond immediate clinical applications, the dataset and results from this study pave the way for the development of artificial intelligence (AI)–driven diagnostic systems. By training machine learning algorithms on validated discriminative parameters, future tools may be able to classify benign and malignant laryngeal disorders with high accuracy using nothing more than a patient's voice sample. Such AI systems could offer real-time, low-cost, and widely accessible diagnostic support, particularly in resource-limited settings where advanced laryngeal imaging modalities are not readily available. The integration of shimmer measures, CPP, LTAS, and other supportive parameters into AI frameworks could revolutionize screening, enabling scalable deployment in primary care, telemedicine, and population-level surveillance.

In summary, this study highlights both robust and exploratory acoustic markers of malignancy, strengthens the evidence for acoustic analysis as a diagnostic adjunct, and outlines the translational potential of these findings in developing future AI-based diagnostic tools. These advances collectively represent a significant step toward more objective, efficient, and widely accessible approaches for the early diagnosis and management of laryngeal cancer.

REFERENCES

1. Fex S, Holmberg EB, Södersten M. Voice characteristics of patients with benign vocal fold lesions. J Voice. 1994;8(3):242–50.

2. Zeitels SM, Hillman RE, Desloge R, et al. Phonomicrosurgery in singers and performing artists. Clin Chest Med. 1991;12(3):553–64.

3. Rabinov CR, Kreiman J, Gerratt BR, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measures. J Speech Lang Hear Res. 1995;38(1):26–32.

4. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. J Acoust Soc Am. 1982;71(6):1544–50.

5. Awan SN, Roy N. Acoustic prediction of voice type in women with functional dysphonia. J Voice. 2005;19(2):268–82.

6. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. J Speech Lang Hear Res. 2001;44(2):327–39.

7. Uloza V, Verikas A, Bacauskiene M. Automated voice analysis applied to laryngeal disease classification. Comput Methods Programs Biomed. 2011;105(3):221–8.

8. De Krom G. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. J Speech Hear Res. 1993;36(2):254–66.

9. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. J Voice. 2004;18(3):299–304.

10. Karnell MP, Scherer RC. Acoustic analysis of laryngeal cancer patients. Laryngoscope. 1991;101(8):857–63.

11. Švec JG, Titze IR. Shimmer in voice: magnitude and frequency components. Folia Phoniatr Logop. 1993;45(5):240–4.

12. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. J Speech Hear Res. 1994;37(4):769–78.

13. Awan SN, Roy N, Zhang D, Cohen SM. Validation of CPP as a clinical tool in dysphonia assessment. J Voice. 2016;30(2):130–4.

14. Lin E, Klatt DH. Spectral analysis of pathological voices. J Acoust Soc Am. 1983;73(3):789–95.