POLITECNICO DI TORINO
Repository ISTITUZIONALE

The Rapidly Evolving Scenario of Acoustic Voice Analysis in Otolaryngology

(Article begins on next page)

08 January 2025

Cureus
Part of SPRINGER NATURE

Open Access Review Article

The Rapidly Evolving Scenario of Acoustic Voice Analysis in Otolaryngology

Marco Fantini [1, 2], Gabriele Ciravegna [3], Alkis Koudounas [3], Tania Cerquitelli [3], Elena Baralis [3], Giovanni Succo [4, 5], Erika Crosetti [4]

1. Ear, Nose, and Throat Unit, Koelliker Hospital, Turin, ITA 2. Ear, Nose, and Throat Unit, San Feliciano Hospital, Rome, Italy 3. DataBase and Data Mining Group (DBDMG) Department of Control and Computer Engineering (DAUIN), Polytechnic University of Turin, Turin, ITA 4. Ear, Nose, and Throat Clinic, Head and Neck Cancer Unit, San Giovanni Bosco Hospital, Turin, ITA 5. Department of Oncology, University of Turin, Turin, ITA

Corresponding author: Marco Fantini, marcofantini8811@hotmail.it

Abstract

The field of voice analysis has experienced significant transformations, evolving from basic perceptual assessments to the incorporation of advanced digital signal processing and computational tools. This progression has facilitated a deeper understanding of the complex dynamics of vocal function, particularly through the use of acoustic voice analysis within a multidimensional evaluation framework. Traditionally, voice analysis relied on parameters such as fundamental frequency, jitter, shimmer, and noise-to-harmonic ratio, which, despite their utility, have faced criticism for variability and lack of robustness. Recent developments have led to a shift toward more reliable metrics such as cepstral measures, which offer improved accuracy in voice quality assessments. Furthermore, the integration of multiparametric constructs underscores a comprehensive approach to evaluating vocal quality, blending sustained vowels, and continuous speech analyses. Current trends in clinical practice increasingly favor these advanced measures over traditional parameters due to their greater reliability and clinical utility. Additionally, the emergence of artificial intelligence (AI), particularly deep learning, holds promise for revolutionizing voice analysis by enhancing diagnostic precision and enabling efficient, non-invasive screening methods. This shift toward AI-driven approaches signifies a potential paradigm change in voice health, suggesting a future where AI not only aids in diagnosis but also the early detection and treatment of voice-related pathologies.

Categories: Preventive Medicine, Otolaryngology
Keywords: acoustic analysis, acoustic voice quality index, artificial intelligence, vocal quality, voice

Introduction And Background

The field of voice analysis has undergone a profound evolution over the last century, transitioning from initial perceptual evaluations to the integration of complex digital signal processing techniques [1-3]. This advancement underscores a pivotal shift in the methodologies employed in phonetics, speech science, and laryngology, aimed at elucidating the complex dynamics of vocal fold vibrations and the acoustic characteristics imparted by the vocal tract's resonance [4-6]. The advent of sophisticated computational tools and software has significantly enhanced the precision and depth of these analyses, supporting a comprehensive understanding of voice quality [7,8].

Acoustic voice analysis is part of a broader, multidimensional approach to voice evaluation, encompassing perceptual voice assessments, self-evaluations, laryngostroboscopic examinations, and acoustic measurements. This holistic framework is essential for a thorough investigation of vocal function, offering insights into the nuanced interplay between physiological mechanisms and acoustic output [9,10].

The clinical utility of acoustic analysis extends beyond mere documentation, serving as a critical component in monitoring treatment progress and facilitating evidence-based therapeutic decisions. By providing objective, quantifiable metrics, it complements perceptual evaluations and laryngostroboscopic findings, enriching the diagnostic process and enhancing the objectivity of treatment outcomes assessment [7,9]. In the research domain, the exploration of acoustic properties has contributed significantly to advancing our knowledge of vocal physiology and the effects of various conditions on voice production [11]. This research has catalyzed the development of innovative therapeutic interventions and preventive strategies, contributing to the optimization of vocal health and performance. In summary, traditional acoustic voice analysis, within the context of a multidimensional evaluation model, embodies the convergence of science and clinical practice in voice care.

Review

Limitations of traditional acoustic parameters for voice analysis

In the landscape of traditional acoustic voice analysis, fundamental frequency (F0), sound pressure level (SPL), and parameters such as jitter, shimmer, and noise-to-harmonic ratio (NHR) have been pivotal [12]. F0

and SPL provide essential insights into pitch and loudness, forming the cornerstone of vocal assessment [3,13]. Conversely, jitter, which measures frequency perturbation, and shimmer, which quantifies amplitude perturbation, alongside NHR, a metric assessing the ratio of noise to harmonic components in the voice, delve into the finer details of voice signal stability, offering nuanced perspectives on the assessment of voice quality and voice disorders [3,12,13].

However, the reliability of these early parameters has been a subject of debate, primarily due to their susceptibility to significant variability, lack of robustness, and sensitivity to extrinsic factors such as frequency and intensity [14,15]. The inherent variability in voice production among individuals, coupled with the influence of recording conditions and analysis methodologies, can introduce a substantial inconsistency in measurements, as demonstrated by Maryn et al. [16]. This variability complicates the interpretation of jitter, shimmer, and NHR values, potentially confounding clinical assessments and research outcomes in different settings [17].

Moreover, the robustness of these microperturbation parameters is further challenged by their dependence on the underlying periodicity of the voice signal. In signals that exhibit a high degree of aperiodicity, common in severe dysphonia or substitution voices, the accuracy and relevance of jitter and shimmer measurements diminish significantly [3,18]. The sensitivity of these parameters to the fundamental frequency and intensity of the voice signal also raises concerns regarding their utility across a wide range of vocal intensities and pitches. Such sensitivity may lead to skewed assessments in cases where vocal parameters deviate from normative ranges, limiting the applicability of these measures in a diverse clinical population.

Another significant methodological limitation affecting jitter, shimmer, and NHR is that their calculation is predominantly based on sustained vowel sounds. This traditional approach encompasses a not negligible variability according to the vowel segment selection, introducing a certain risk of inconsistency across measurements and individuals [19]. Moreover, while methodically simpler, vowel analysis encounters limitations in ecological validity, as real-world voice use encompasses a dynamic range of vocal expressions beyond sustained vowels. The analysis of these parameters on sustained vowels, although informative, does not fully encapsulate the complexities of prosody [20]. On the other hand, in the context of continuous speech, the vocal signal presents a higher degree of variability and complexity, leading to potential inaccuracies in microperturbation measurements. This reduction in reliability stems from the parameters' sensitivity to the periodicity of the voice signal, which is more pronounced in sustained vowels than in the variable pitch and intensity landscape of continuous speech.

In light of these limitations, the reliance on traditional parameters necessitates a cautious interpretation within the context of a comprehensive voice assessment. The recognition of these constraints has propelled the field toward the development and adoption of more robust, reliable, and multiparametric acoustic measures.

Cepstral measures for voice analysis

Cepstral indices, particularly the cepstral peak prominence (CPP) and its smoothed variant, CPPs (smoothed cepstral peak prominence), represent a significant advancement in the acoustic analysis of voice [21]. These indices are derived from a mathematical representation of the voice signal known as the cepstrum, which is the result of taking the Fourier transform of the logarithm of the power spectrum of the voice signal [22], as shown in Figure 1. This process effectively captures and quantifies the periodicity of the vocal signal, offering insights into the harmonic structure and the presence of noise within the voice.
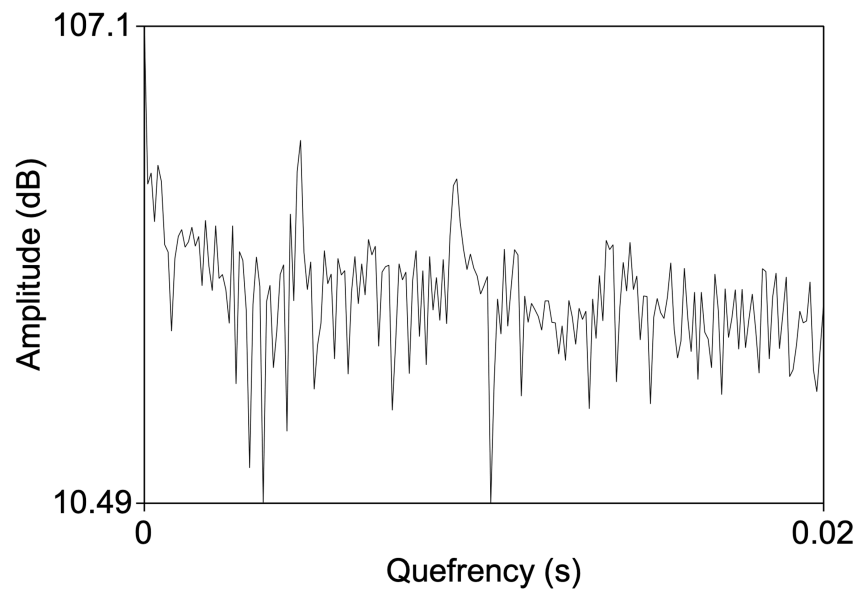
2024 Fantini et al. Cureus 16(11): e73491. DOI 10.7759/cureus.73491

2 of 8

FIGURE 1: Graphic representation of a cepstrum.

This figure was obtained using the free software PRAAT (www.fon.hum.uva.nl/praat/) by the author Marco Fantini.

The CPP, and especially CPPs, have gained prominence due to their ability to provide a reliable, robust, and reproducible measure of voice quality, particularly in the assessment of dysphonia [23]. Unlike traditional parameters such as jitter and shimmer, cepstral measures do not require the voice signal to be perfectly periodic, making them more applicable across a wider range of voice qualities, including those with significant aperiodicity often observed in severely dysphonic voices [24,25]. Moreover, cepstral measures calculation is less susceptible to variations in recording conditions or the specific characteristics of the analyzed voice sample, leading to more consistent and reproducible results across different testing environments and populations [26].

The increasing utilization and trust in cepstral measures stem from their strong correlation with perceptual assessments of voice quality. Research has demonstrated that CPPs, in particular, is effective in distinguishing between normal and dysphonic voices both on sustained vowels and continuous speech, correlating well with the severity of vocal dysfunction [27]. This correlation is crucial in clinical settings where objective measures must align closely with perceptual evaluations to guide diagnosis and treatment. In summary, cepstral measures, particularly CPPs, have emerged in the last years as pivotal tools in the acoustic analysis of voice, providing a detailed and accurate assessment of voice quality and dysphonia [23,24].

Multiparametric constructs for voice analysis: a focus on the Acoustic Voice Quality Index

In the advancing field of voice analysis, the development of multiparametric constructs marks a significant stride toward a more holistic assessment of vocal quality. These constructs integrate various acoustic parameters to provide a comprehensive assessment of voice characteristics. Among the notable constructs are the Acoustic Voice Quality Index (AVQI) [28], the Cepstral Spectral Index of Dysphonia (CSID), and the Acoustic Breathiness Index (ABI) [29], which amalgamate multiple acoustic measures to evaluate specific aspects of voice quality, such as overall vocal quality and breathiness.

A significant advancement in this realm is undoubtedly the AVQI, proposed by Maryn et al. [28]. The AVQI stands out as a robust, multivariable acoustic tool designed to quantify auditorily perceived overall voice quality by analyzing both sustained vowel sounds and voiced fragments of continuous speech [16]. It is calculated through a specific script run by the software PRAAT and is displayed as shown in Figure *2*. Since its inception in 2010, AVQI has undergone refinements [30] and validation processes across several different languages worldwide, highlighting its universal applicability and reliability in voice quality assessment [31-39]. AVQI's formulation includes a variety of acoustic measures, such as the HNR, shimmer, and CPPs, among others [28,30]. The AVQI emerges as a notable ecological measure of voice quality within the domain of acoustic analysis, principally due to its comprehensive approach that encompasses concatenated sustained vowels and continuous speech. This integration represents a significant advancement over traditional acoustic analyses, which predominantly focus on sustained vowels. The inclusion of continuous

2024 Fantini et al. Cureus 16(11): e73491. DOI 10.7759/cureus.73491

3 of 8

speech in AVQI's assessment framework addresses a critical limitation of earlier methods, offering a more representative and holistic evaluation of vocal function as encountered in daily life.
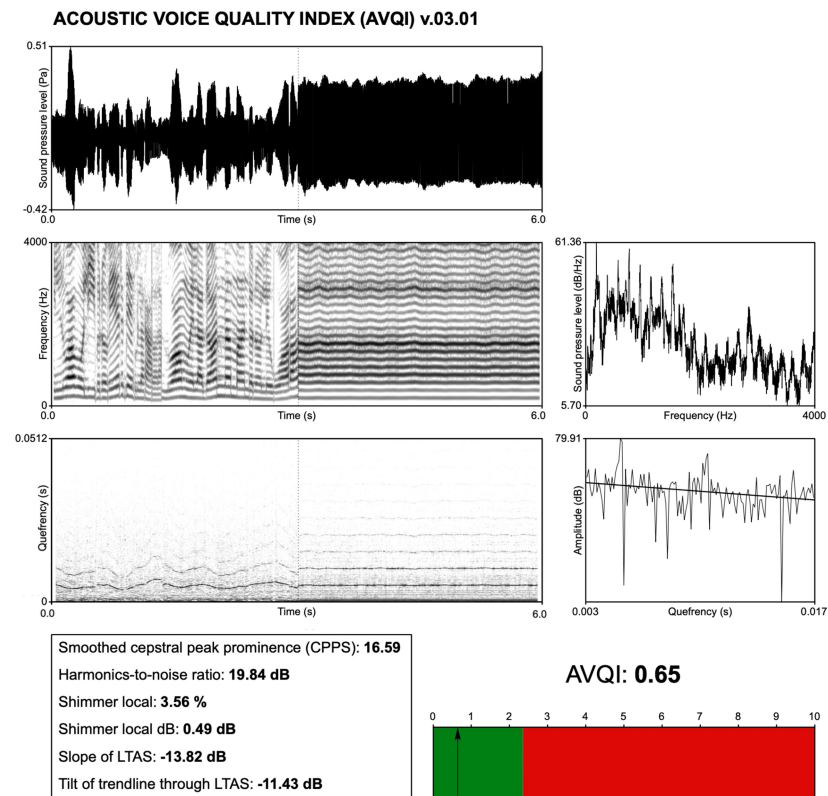
ACOUSTIC VOICE QUALITY INDEX (AVQI) v.03.01

Smoothed cepstral peak prominence (CPPS): **16.59**

Harmonics-to-noise ratio: **19.84 dB**

Shimmer local: **3.56 %**

Shimmer local dB: **0.49 dB**

Slope of LTAS: **-13.82 dB**

Tilt of trendline through LTAS: **-11.43 dB**

AVQI: **0.65**

FIGURE 2: Example of the Acoustic Voice Quality Index (AVQI), version 03.01, with the Italian language cut-off 2.35.

This figure was obtained using the free software PRAAT (www.fon.hum.uva.nl/praat/) by the author Marco Fantini.

The comprehensive validation and widespread adoption of AVQI were highlighted in a recent meta-analysis, which drew attention to its strong diagnostic accuracy and clinical utility [40]. The meta-analysis reported an area under the curve (AUC) of 0.937 and a Q* index of 0.874, demonstrating AVQI's exceptional capability in diagnosing voice disorders. Furthermore, the analysis underscored AVQI's concurrent validity and sensitivity to change, with weighted mean correlation coefficients of 0.838 and 0.796, respectively. These findings affirm the AVQI's effectiveness in mirroring perceptual evaluations of voice quality and tracking changes over time, whether due to natural progression, surgical intervention, or therapy.

Current recommendations on acoustic voice analysis: where are we now?

Recent clinical recommendations on voice acoustic analysis are progressively advocating for the use of cepstral measures or multiparametric indices for voice analysis. Authoritative examples include the protocols proposed by the American Speech-Language-Hearing Association (ASHA) [8] and those by Behlau et al. in Brazil [41]. The first paper suggests CPP, the latter CPPs, AVQI, and CSID as acoustic voice quality parameters. These recommendations reflect a growing recognition of the limitations of traditional acoustic parameters and emphasize the superior reliability, robustness, and clinical utility of cepstral and multiparametric measures in assessing vocal function and disorders.

In contrast, the recent European guidelines, emerging from an expert consensus panel of the European Laryngological Society and Union of European Phoniatricians, continue to recommend the use of traditional measures such as jitter, shimmer, and NHR for voice analysis in clinical settings [42]. These guidelines notably lack mention of cepstral parameters and offer scant citations of multiparametric constructs despite mounting evidence supporting their efficacy in voice analysis. This discrepancy underscores a significant divergence in the adoption of advanced voice analysis techniques across different regions and expert groups. The persistent recommendation of traditional measures, which are increasingly recognized for their

limitations in ecological validity and sensitivity, highlights the need for an updated consensus that integrates contemporary findings and methodologies.

The adoption of cepstral measures and multiparametric indices in voice analysis represents a significant advancement in our ability to accurately and comprehensively assess voice quality and voice disorders. These measures offer a more nuanced understanding of vocal function, facilitating effective diagnosis, monitoring, and treatment planning. Therefore, the integration of modern acoustic analysis techniques into international standards would not only enhance the accuracy and reliability of voice assessments but also promote consistency in voice care practices internationally, ultimately benefiting patients and clinicians alike.

Toward the future: artificial intelligence for voice analysis and voice restoration

The last decades witnessed an increasing recognition within the medical scientific community about the significance of artificial intelligence (AI) in healthcare. This growing awareness is driven by emerging evidence showcasing AI's potential to revolutionize various aspects of health-related research and clinical practice. Deep learning (DL) is a specialized domain within AI and machine learning (ML). AI encompasses algorithms tailored for tasks that traditionally require human-like intelligence. A distinctive feature of DL is its capacity to autonomously identify significant features from training data, eliminating the need for pre-defined feature specifications. The foundation of most DL models is convolutional neural networks and recurrent neural networks, which draw methodological parallels to biological neural networks. These networks consist of multiple strata of interlinked neurons, where each neuron functions as a simple mathematical function, processing and weight-averaging the outputs from preceding layers. Data input is fed into the model, leading to the generation of the model's prediction. This output is then compared against actual values, facilitating the computation of errors. The network adjusts its synaptic weights (connections between neurons) based on these errors. Connections pivotal to accurate predictions are strengthened, whereas less significant ones are assigned lower weights. The network ultimately calculates a probability for the given input. For instance, in a healthcare application, the input might be a patient's voice recording, and the network might output a binary classification, such as distinguishing between a normal voice and a pathological voice [43].

DL has emerged as a transformative force across various healthcare domains, ranging from the analysis of electronic health records [44] to medical imaging [45], drug discovery [46], and multi-omics analysis [47], achieving accuracies comparable to, or surpassing, those of human experts. Considering voice analysis, DL is gaining traction in this field as well, with studies demonstrating its potential to revolutionize the screening and early diagnosis of voice pathologies and voice-related diseases (such as neurological disorders) [48], offering a non-invasive, efficient alternative to traditional diagnostic methods. A recent systematic review by Tessler et al. analyzed 14 studies on the application of AI in voice analysis, showcasing promising diagnostic accuracy. This highlights AI's potential for early diagnosis and screening using vocal signals, including remote applications [43].

A step forward might be represented by the application of the so-called Transformer models to voice analysis [49]. Leveraging the self-attention mechanism, Transformers can discern patterns by comparing different pieces of data (tokens) even when widely spaced. This capability enables them to understand the context of each token effectively [50], making them suitable for analyzing voice data and capturing subtle nuances in speech patterns and intonations. Their parallel processing capability allows for the efficient handling of sequential information such as language. Initial studies [51,52] suggest that Transformers significantly improve voice-based diagnostics by identifying vocal features associated with specific pathologies. Furthermore, these models excel in handling diverse modalities simultaneously, promising a holistic understanding of an individual's health status [53].

Apart from voice analysis and early diagnosis of voice-related pathologies, emerging applications of AI for voice restoration, particularly for patients with severely deteriorated voices due to conditions such as laryngectomies or cordectomies for laryngeal cancers [54] or neurological disorders [48], present intriguing possibilities. The impact of these conditions on voice quality significantly affects patients' voice-related quality of life, especially in crucial areas such as telephone communication [55]. AI could play a key role in this context, fostering the development of systems aimed at real-time audio signal manipulation to restore clear and intelligible speech, both through smartphone applications [56] and wearable devices [57]. This advancement would greatly enhance communication abilities and quality of life for affected individuals.

Conclusions

In the rapidly evolving scenario of voice acoustic analysis, there has been a significant progression from basic techniques, which were subject to high variability and limited reliability, to increasingly complex constructs and more robust methods. These advanced techniques are gradually overcoming the limitations inherent in wave perturbation measures, marking a notable evolution in the field. Moreover, the last years witnessed a clear trend toward the adoption of AI for voice analysis. This evolution signifies a potential shift from traditional methods to more advanced, AI-driven techniques. AI's promise lies in its capacity for

remote, efficient, and precise screening of voice-related pathologies, likely shaping the future of this field. This development could revolutionize the approach to voice health, offering new possibilities for screening, early detection, and intervention in vocal disorders.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Marco Fantini, Gabriele Ciravegna, Alkis Koudounas, Tania Cerquitelli, Elena Baralis, Giovanni Succo, Erika Crosetti

Drafting of the manuscript: Marco Fantini, Gabriele Ciravegna, Alkis Koudounas

Critical review of the manuscript for important intellectual content: Tania Cerquitelli, Elena Baralis, Giovanni Succo, Erika Crosetti

Supervision: Tania Cerquitelli, Elena Baralis, Giovanni Succo

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Hirano M: Psycho-acoustic evaluation of voice . Clinical Examination of Voice. Hirano M (ed): Springer-Verlag, New York; 1981. 81-4.
2. Dejonckere PH, Obbens C, de Moor GM, Wieneke GH: Perceptual evaluation of dysphonia: reliability and relevance. Folia Phoniatr (Basel). 1993, 45:76-83. 10.1159/000266220
3. Titze IR: Workshop on Acoustic Analysis: Summary Statement . National Center for Voice and Speech, Iowa City, LA; 1995.
4. Titze IR: A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice . J Voice. 2004, 18:292-8. 10.1016/j.jvoice.2003.12.010
5. Sundberg J: The acoustics of the singing voice . Sci Am. 1977, 236:82-4, 86, 88-91. 10.1038/scientificamerican0377-82
6. Ternström S, Pabon P, Södersten M: The voice range profile: its function, applications, pitfalls and potential. Acta Acust United Acust. 2016, 102:268-83. 10.3813/AAA.918943
7. Boersma P, Weenink D: Praat: Doing Phonetics by Computer . Praat, Amsterdam; 2024.
8. Patel RR, Awan SN, Barkmeier-Kraemer J, et al.: Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. Am J Speech Lang Pathol. 2018, 27:887-905. 10.1044/2018_AJSLP-17-0009
9. Dejonckere PH, Bradley P, Clemente P, et al.: A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). Eur Arch Otorhinolaryngol. 2001, 258:77-82. 10.1007/s004050000299
10. Rusz J, Tykalova T, Ramig LO, Tripoliti E: Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. Mov Disord. 2021, 36:803-14. 10.1002/mds.28465
11. Brockmann-Bauser M, de Paula Soares MF: Do we get what we need from clinical acoustic voice measurements?. Appl Sci. 2023, 13:941. 10.3390/app13020941
12. Teixeira JP, Oliveira C, Lopes C: Vocal acoustic analysis - jitter, shimmer and HNR parameters . Proc Technol. 2013, 9:1112-22. 10.1016/j.protcy.2013.12.124
13. Ricci Maccarini A, De Colle W, Lucchini E, Casolino D: L'esame spettroacustico della voce. Le Disfonie: Fisiopatologia, Clinica ed Aspetti Medico-legali, Relazione Ufficiale. LXXXIX SIO National Congress, Italy; 2002. 150-82.
14. Brockmann-Bauser M, Bohlender JE, Mehta DD: Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders. J Voice. 2018, 32:162-8. 10.1016/j.jvoice.2017.04.008
15. de Oliveira Florencio V, Almeida AA, Balata P, Nascimento S, Brockmann-Bauser M, Lopes LW: Differences and reliability of linear and nonlinear acoustic measures as a function of vocal intensity in individuals with voice disorders. J Voice. 2023, 37:663-81. 10.1016/j.jvoice.2021.04.011
16. Maryn Y, Corthals P, De Bodt M, Van Cauwenberge P, Deliyski D: Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat. Folia Phoniatr Logop. 2009, 61:217-26. 10.1159/000227999
17. Karnell MP, Hall KD, Landahl KL: Comparison of fundamental frequency and perturbation measurements among three analysis systems. J Voice. 1995, 9:383-93. 10.1016/s0892-1997(05)80200-0

18. Sprecher A, Olszewski A, Jiang JJ, Zhang Y: Updating signal typing in voice: addition of type 4 signals . J Acoust Soc Am. 2010, 127:3710-16. 10.1121/1.3397477
19. Choi SH, Lee J, Sprecher AJ, Jiang JJ: The effect of segment selection on acoustic analysis . J Voice. 2012, 26:1-7. 10.1016/j.jvoice.2010.10.009
20. Zraick RI, Wendel K, Smith-Olinde L: The effect of speaking task on perceptual judgment of the severity of dysphonic voice. J Voice. 2005, 19:574-81. 10.1016/j.jvoice.2004.08.009
21. Hillenbrand J, Houde RA: Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. J Speech Hear Res. 1996, 39:311-21. 10.1044/jshr.3902.311
22. Ferrer Riesgo CA, Nöth E: What makes the cepstral peak prominence different to other acoustic correlates of vocal quality?. J Voice. 2020, 34:806.e1-6. 10.1016/j.jvoice.2019.01.004
23. Aghajanzadeh M, Saeedi S: Efficacy of cepstral measures in voice disorder diagnosis: a literature review . J Mod Rehabil. 2022, 16:120-9. 10.18502/jmr.v16i2.9298
24. Heman-Ackah YD, Heuer RJ, Michael DD, et al.: Cepstral peak prominence: a more reliable measure of dysphonia. Ann Otol Rhinol Laryngol. 2003, 112:324-33. 10.1177/000348940311200406
25. Heman-Ackah YD, Michael DD, Goding GS Jr: The relationship between cepstral peak prominence and selected parameters of dysphonia. J Voice. 2002, 16:20-7. 10.1016/s0892-1997(02)00067-x
26. Leong K, Hawkshaw MJ, Dentchev D, Gupta R, Lurie D, Sataloff RT: Reliability of objective voice measures of normal speaking voices. J Voice. 2013, 27:170-6. 10.1016/j.jvoice.2012.07.005
27. Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P: Acoustic measurement of overall voice quality: a meta-analysis. J Acoust Soc Am. 2009, 126:2619-34. 10.1121/1.3224706
28. Maryn Y, De Bodt M, Roy N: The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. J Commun Disord. 2010, 43:161-74. 10.1016/j.jcomdis.2009.12.004
29. Barsties V Latoszek B, Mathmann P, Neumann K: The cepstral spectral index of dysphonia, the acoustic voice quality index and the acoustic breathiness index as novel multiparametric indices for acoustic assessment of voice quality. Curr Opin Otolaryngol Head Neck Surg. 2021, 29:451-7. 10.1097/MOO.0000000000000743
30. Barsties B, Maryn Y: External validation of the Acoustic Voice Quality Index Version 03.01 with extended representativity. Ann Otol Rhinol Laryngol. 2016, 125:571-83. 10.1177/0003489416636131
31. Hosokawa K, Barsties B, Iwahashi T, et al.: Validation of the Acoustic Voice Quality Index in the Japanese language. J Voice. 2017, 31:260.e1-9. 10.1016/j.jvoice.2016.05.010
32. Zainaee S, Khadivi E, Jamali J, Sobhani-Rad D, Maryn Y, Ghaemi H: The acoustic voice quality index, version 2.06 and 3.01, for the Persian-speaking population. J Commun Disord. 2022, 100:106279. 10.1016/j.jcomdis.2022.106279
33. Barsties V Latoszek B, Lehnert B, Janotte B: Validation of the Acoustic Voice Quality Index Version 03.01 and Acoustic Breathiness Index in German. J Voice. 2020, 34:157.e17-25. 10.1016/j.jvoice.2018.07.026
34. Pommée T, Maryn Y, Finck C, Morsomme D: Validation of the Acoustic Voice Quality Index, Version 03.01, in French. J Voice. 2020, 34:646.e11-26. 10.1016/j.jvoice.2018.12.008
35. Kankare E, Barsties V Latoszek B, Maryn Y, et al.: The acoustic voice quality index version 02.02 in the Finnish-speaking population. Logoped Phoniatr Vocol. 2020, 45:49-56. 10.1080/14015439.2018.1556332
36. Englert M, Barsties V Latoszek B, Maryn Y, Behlau M: Validation of the Acoustic Voice Quality Index, Version 03.01, to the Brazilian Portuguese language. J Voice. 2021, 35:160.e15-21. 10.1016/j.jvoice.2019.07.024
37. Kim GH, von Latoszek BB, Lee YW: Validation of Acoustic Voice Quality Index Version 3.01 and Acoustic Breathiness Index in Korean population. J Voice. 2021, 35:660.e9-660.e18. 10.1016/j.jvoice.2019.10.005
38. Yeşilli-Puzella G, Tadıhan-Özkan E, Maryn Y: Validation and test-retest reliability of Acoustic Voice Quality Index Version 02.06 in the Turkish language. J Voice. 2022, 36:736.e25-32. 10.1016/j.jvoice.2020.08.021
39. Fantini M, Ricci Maccarini A, Firino A, et al.: Validation of the Acoustic Voice Quality Index (AVQI) Version 03.01 in Italian. J Voice. 2023, 37:631.e1-6. 10.1016/j.jvoice.2021.02.029
40. Batthyany C, Latoszek BB, Maryn Y: Meta-Analysis on the validity of the Acoustic Voice Quality Index [in press]. J Voice. 2022, 10.1016/j.jvoice.2022.04.022
41. Behlau M, Madazio G, Vaiano T, Pacheco C, Badaró F: Voice evaluation - contribution of the speech-language pathologist voice specialist - SLP-V: part B. Acoustic analysis, physical examination and correlation of all steps with the medical diagnoses. Hear Balance Commun. 2021, 19:318-24. 10.1080/21695717.2022.2028490
42. Lechien JR, Geneid A, Bohlender JE, et al.: Consensus for voice quality assessment in clinical practice: guidelines of the European Laryngological Society and Union of the European Phoniatricians. Eur Arch Otorhinolaryngol. 2023, 280:5459-73. 10.1007/s00405-023-08211-6
43. Tessler I, Primov-Fever A, Soffer S, et al.: Deep learning in voice analysis for diagnosing vocal cord pathologies: a systematic review. Eur Arch Otorhinolaryngol. 2024, 281:863-71. 10.1007/s00405-023-08362-6
44. Rajkomar A, Oren E, Chen K, et al.: Scalable and accurate deep learning with electronic health records . NPJ Digit Med. 2018, 1:18. 10.1038/s41746-018-0029-1
45. Suzuki K: Overview of deep learning in medical imaging . Radiol Phys Technol. 2017, 10:257-73. 10.1007/s12194-017-0406-5
46. Jumper J, Evans R, Pritzel A, et al.: Highly accurate protein structure prediction with AlphaFold. Nature. 2021, 596:583-9. 10.1038/s41586-021-03819-2
47. Lovino M, Randazzo V, Ciravegna G, Barbiero P, Ficarra E, Cirrincione G: A survey on data integration for multi-omics sample clustering. Neurocomputing. 2022, 488:494-508. 10.1016/j.neucom.2021.11.094
48. Islam MA, Hasan Majumder MZ, Hussein MA, Hossain KM, Miah MS: A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. Heliyon. 2024, 10:e25469. 10.1016/j.heliyon.2024.e25469
49. Vaswani A, Shazeer N, Parmar N, et al.: Attention is all you need . 31st Conference on Neural Information Processing Systems. 2017, 1-11.
50. Ribas D, Pastor MA, Miguel A, Martınez D, Ortega A, Lleida E: Automatic voice disorder detection using

2024 Fantini et al. Cureus 16(11): e73491. DOI 10.7759/cureus.73491

7 of 8

self-supervised representations. IEEE Acc. 2023, 11:14915-27. 10.1109/ACCESS.2023.3243986

51. La Quatra M, Vaiani L, Koudounas A, Cagliero L, Garza P, Baralis E: How much attention should we pay to mosquitoes?. Proceedings of the 30th ACM International Conference on Multimedia. 2022, 7135-9. 10.1145/3503161.3551594

52. Koudounas A, Ciravegna G, Fantini M, Succo G, Crosetti E, Cerquitelli T, Baralis E: Voice disorder analysis: a transformer-based approach. arXiv. 2024, 10.21437/Interspeech.2024-1122

53. Reed S, Zolna K, Parisotto E, et al.: A generalist agent. arXiv. 2022, 10.48550/arXiv.2205.06175

54. Fantini M, Crosetti E, Pizzorni N, et al.: Voice and communication after open partial horizontal laryngectomies: a cross-sectional outcome study. Head Neck. 2022, 44:2248-56. 10.1002/hed.27132

55. Crosetti E, Fantini M, Arrigoni G, et al.: Telephonic voice intelligibility after laryngeal cancer treatment: is therapeutic approach significant?. Eur Arch Otorhinolaryngol. 2017, 274:337-46. 10.1007/s00405-016-4217-9

56. Whispp App. Assistive Voice Technology for people with a voice disability and people who stutter severely . (2024). http://www.whispp.com.

57. Che Z, Wan X, Xu J, Duan C, Zheng T, Chen J: Speaking without vocal folds using a machine-learning-assisted wearable sensing-actuation system. Nat Commun. 2024, 15:1873. 10.1038/s41467-024-45915-7

2024 Fantini et al. Cureus 16(11): e73491. DOI 10.7759/cureus.73491

8 of 8