Consensus-Based Definitions for Vocal Biomarkers: The International VOCAL Initiative

Author list: Mégane Pizzimenti, PhD[1]; Ayush Kalia, BSc[2,3]; Jamie A. Toghranegar, SLPD[2]; Mohamed Ebraheem, MSc[2]; Nicholas Cummings, PhD[4]; Satrajit S. Ghosh, PhD[5]; James T. Anibal, BSc[6,7]; Rhoda Au, PhD[8,9,10,11,12,13]; Arian Azarang, PhD[14]; Ruth H. Bahr, PhD[15]; Steven D. Bedrick, PhD[16]; Hugo Botha, MB, ChB[17]; Oita C. Coleman, BSc[18]; Abir Elbeji, PhD[1]; Lampros C. Kourtis, PhD[19]; Anaïs Rameau, MD, MSc[20]; Jaskanwal Deep Singh Sara, MD[21,22]; Stephanie W. Watts, PhD[2]; Daria Hemmerling, PhD[23]; Jiri Mekyska, PhD[24,25]; Marisha L. Speights, PhD[26]; Jean-Christophe Bélisle-Pipon, PhD[27]; Yael E. Bensoussan*, MD[2]*; Guy Fagherazzi, PhD[1]*
* Contributed equally

Affiliations:

1. Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health, Strassen, Luxembourg.
2. USF Health Voice Center, Department of Otolaryngology-Head & Neck Surgery, University of South Florida, Tampa, FL, United States.
3. Department of Internal Medicine, Yale School of Medicine, New Haven, United States.
4. Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, United Kingdom.
5. McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States.
6. Center for Interventional Oncology, NIH Clinical Center, National Institutes of Health, Bethesda, United States.
7. Computational Health Informatics Lab, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom.
8. Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, United States.
9. Department of Anatomy & Neurobiology, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, United States.
10. Boston University Alzheimer's Disease Research Center, Boston, Massachusetts, United States.
11. The Framingham Heart Study, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, United States.
12. Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, United States.
13. Department of Neurology, Boston University Chobanian & Avedisian School of Medicine,

Boston, Massachusetts, United States.
14. Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States.
15. Department of Communication Sciences & Disorders, University of South Florida, Tampa, FL, United States.
16. Departments of Medical Informatics and Clinical Epidemiology and Medicine, Oregon Health and Science University, Portland, OR, United States.
17. Department of Neurology, Mayo Clinic, Rochester, MN, United States.
18. Open Voice TrustMark Initiative, Apex, NC, United States.
19. Alzheimer's Drug Discovery Foundation, Boston, MA, United States.
20. Sean Parker Institute for the Voice, Department of Otolaryngology-Head & Neck Surgery, Weill Cornell Medical College, New York, NY, United States.
21. Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, United States.
22. Minneapolis Heart Institute and Minneapolis Heart Institute Foundation, Abbott Northwestern Hospital, Minneapolis, Minnesota, United States.
23. Department of Measurement and Electronics, AGH University of Krakow, Kraków, 30-059, Poland.
24. Department of Telecommunications, Brno University of Technology, Brno, Czech Republic.
25. Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic.
26. Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60202, United States.
27. Department of Bio-ethics, Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada.

Corresponding author: Guy Fagherazzi, PhD, Luxembourg Institute of Health, Department of Precision Health, 1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg, Guy.Fagherazzi@lih.lu.

Word count: 3374

Key points

Question: Can we establish an international consensus on the definitions related to vocal biomarkers?

Findings: Through a multi-stage international consensus process involving expert representatives from the Bridge2AI-Voice Consortium and eVoiceNet Network, a hierarchical model of vocal biomarkers was developed, defining terms from broad concepts (biomarker, digital biomarker, vocal biomarker) to specific physiological and cognitive domains (cardio-respiratory acoustic, voice, speech/articulatory, cognitive/language).

Meaning: Standardized definitions for vocal biomarkers provide an essential shared vocabulary for interdisciplinary collaboration and lay the foundational groundwork for future guidelines needed to advance the implementation of voice-based health technologies into clinical practice and clinical research.

Abstract

Importance: Voice-based health technologies are growing rapidly, but they lack standardized terminology, which hinders interdisciplinary collaboration, research quality, and clinical translation.

Objective: The objective of this work is to develop universally accepted definitions in the rapidly evolving field of vocal biomarkers, as part of the VOCAL (Vocal Biomarker Guidelines for Ontology, Classification, Application, and Logistics) initiative, a structured, international consensus-based framework that aims to provide standards, and guidelines.

Design: VOCAL is a rigorous, international, multi-stage consensus-building study conducted in 2024-2025.

Setting: Multi-institutional collaboration between representatives from the Bridge2AI-Voice Consortium (North America) and the eVoiceNet Network (European Union), culminating in an in-person workshop at the 2025 Bridge2AI Voice Symposium.

Participants: A group of 24 international experts in medicine, clinical research, speech and language, audio signal processing, statistics, methodology, regulation, ethics.

Methods: VOCAL's iterative process involved five rounds of review, feedback, and an in-person workshop at the international 2025 Bridge2AI Voice Symposium, ensuring the incorporation of diverse perspectives and achieving a robust agreement on the proposed definitions.

Main outcomes and measures: Consensus-based definitions for vocal biomarkers, spanning

from broad concepts (biomarker, digital biomarker, vocal biomarker) to domain-specific measures (cardio-respiratory acoustic, voice, speech/articulatory, cognitive/language).

Results: A hierarchical continuum model of vocal biomarkers was established. We first distinguished between the concepts of vocal measures and vocal biomarkers. We then defined terms from broad, overarching concepts (Level 0: Biomarker, Digital Biomarker, Vocal Biomarker) to more specific physiological and cognitive domains (Level 1: Cardio-Respiratory Acoustic; Level 2: Voice; Level 3: Speech/Articulatory; Level 4: Cognitive/Language, including linguistic and paralinguistic subtypes).

Conclusions and Relevance: This work provides a shared vocabulary that is essential for fostering communication through interdisciplinary collaboration, improving the quality and efficiency of research and development, and ensuring the ethical, reliable, and scalable deployment of future voice-based health technologies. It lays foundational groundwork for upcoming guidelines and standards, which are crucial for advancing the field of vocal biomarkers into widespread clinical utility.

Keywords

Digital Health; Digital Voice; Digital Speech; Biomarker; Digital Biomarker; Vocal Biomarker; Voice Biomarker; Speech Biomarker; AI; Signal Processing

Introduction

In recent years, the concept of vocal biomarkers has garnered considerable attention for its transformative potential in detecting, monitoring, and managing various health conditions through the analysis of voice-based technologies.[1,2] Vocal biomarkers fall under the broader category of digital biomarkers, which are objective and quantifiable health indicators derived from data collected via digital devices such as wearables and sensors. These non-invasive digital biomarkers promise to revolutionize clinical trials and healthcare delivery by enabling continuous, objective, and patient-centered data collection.[3]

A key barrier to making vocal biomarkers a reality is the presumption that any digitally acquired voice metric is a vocal biomarker, despite not having undergone rigorous validation. Making a parallel with imaging and blood-based biomarkers, extensive research and investment over many years have gone into validating Alzheimer's disease imaging and blood-based measures as biomarkers. This same careful, scientific approach for biomarker validation is not currently being applied to digital data. When validating any biomarker, the context of use must be defined to establish it as a biomarker for a specific disease. Thus, moving the field from measuring features of the voice and speech to validated vocal biomarkers will require the time, energy, and investment necessary to undergo validation processes that biological biomarkers have already undergone. The path to validated biomarkers is generally defined by authorities such as the Food and Drug Administration (FDA) or the European Medicines Agency (EMA). Guidelines for digital biomarkers from these regulatory agencies are still in their nascent stages. Thus, there is an opportunity to collaborate with them in developing validation pipelines for various types of vocal biomarkers. We provide definitions of different types of vocal biomarkers below, with the caveat that validation is needed for each subtype within each category and for a specific context of use. Future work will require a transparent validation process and address concerns

regarding artificial intelligence (AI)-driven approaches to vocal biomarker identification, details of which are discussed below.

Once validated, vocal biomarkers will be unique among digital biomarkers in that they derive from the interplay of multiple physiological systems (respiratory, phonatory, articulatory, and cognitive) captured through audio. Their integration into healthcare reflects a broader movement toward more real-time and remote health sensing.[4,5] In Parkinson's disease, speech and language impairments are common and often precede motor dysfunction; features such as reduced loudness, monotone speech, imprecise articulation, and altered rhythm show strong potential for early detection and tracking disease severity and progression.[6] In Alzheimer's disease, lexical-semantic and acoustic measures show high diagnostic performance for mild cognitive impairment, correlate with amyloid-β status and hippocampal volume, and predict 2-year disease progression, supporting their use for early detection and monitoring.[7] Psychiatric and neurodevelopmental conditions likewise exhibit distinctive vocal patterns: for example, depressed individuals tend to speak more slowly, quietly, and with flattened prosody (with frequent pauses), and these acoustic characteristics correlate with depression severity.[8] In chronic heart failure, voice analysis offers a promising, cost-effective, and convenient alternative for diagnosing, risk prediction, and telemonitoring clinical outcomes.[9] Vocal biomarkers also help identify smoking status from ecological audio recordings,[10] and show promise as a screening or monitoring tool for conditions like type 2 diabetes.[11]

Despite this rapid growth, a fundamental challenge persists: the absence of universally accepted definitions for key terms such as "voice biomarkers", "vocal biomarkers", and "speech biomarkers", among others. These terms are frequently used interchangeably in the literature,[10,12–18], yet they reflect features originating from distinct physiological and cognitive systems. A review of existing research reveals a wide variance in definitions, often leading to

nearly as many interpretations as there are articles on the topic.[4] Such terminological ambiguity presents a significant barrier to scientific progress.[19] It hampers collaboration across diverse disciplines, complicates efforts to share data, and impedes the development of standardized methodologies and best practices.[20] Without a common language, the field risks fragmentation and inefficiency, ultimately undermining its scientific advancement, regulatory assessment, clinical translation, and trustworthiness of its innovation.[21,22]

This challenge is further reinforced by a set of interrelated risks that threaten the field's coherence and scalability. The widespread use of AI, proprietary algorithms, and closed datasets has created a "black box" problem, where the lack of transparency limits reproducibility and erodes trust among researchers, clinicians, and regulators. In parallel, inconsistent terminologies and disciplinary silos exacerbate confusion, highlighting the urgent need for a shared, evolving vocabulary. The technical landscape adds further complexity, as voice data is often massive, heterogeneous, and collected under diverse conditions, making integration across studies difficult without adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles.[23] Voice data-specific regulatory frameworks remain fragmented and difficult to navigate, particularly across regions, posing significant hurdles to clinical validation and implementation.[24] Additionally, concerns about privacy, data ownership, and ethical use of sensitive voice data persist,[25] alongside practical challenges such as ensuring sustained patient engagement in digital trials.[26] A recent survey of stakeholders in voice AI development emphasized the need for ethically sourced, diverse, and transparent data practices, calling for trustworthiness not just as a technical feature but as a guiding norm.[25]

The lack of consensus also impairs the development of methodological infrastructure. While master protocols are now standard in clinical trials and other biomarker fields, vocal biomarker research lacks unified frameworks, resulting in heterogeneous data collection, feature

extraction, and analysis methods.[4] This absence complicates validation and clinical scalability. These converging barriers underscore the pressing need for community-driven standards, transparent methodologies, and harmonized regulatory pathways to unlock the full potential of vocal biomarkers. Establishing a unified framework for vocal biomarkers is therefore essential. Standardization will provide a shared vocabulary that facilitates interdisciplinary collaboration, supports data interoperability, and enables reproducibility of research findings. This foundation is critical for advancing vocal biomarker research from exploratory studies to reliable clinical applications.

To address this challenge, this paper proposes a set of foundational definitions as the first step of the VOCAL initiative (standing for Vocal Biomarker Guidelines for Ontology, Classification, Application, and Logistics), a structured, international framework for harmonizing practices and guiding the field of vocal biomarkers. We aim to develop a shared and controlled vocabulary that promotes clarity and consistency across disciplines. This will facilitate effective communication among all stakeholders and the general public, and it will guide the development of best practices for the field.

Methods

Consensus-Building Process

To develop clear and widely applicable definitions of vocal biomarkers, a structured, multi-stage consensus-building process was employed. This approach was meticulously designed to incorporate diverse perspectives from interdisciplinary stakeholders, ensuring transparency and rigor throughout the development of the framework. The iterative nature of the process enabled continuous refinement and validation of the proposed definitions, resulting in a robust and

widely accepted outcome.

Participants

The consensus process involved three distinct groups of participants, ensuring broad representation and a comprehensive range of expertise:

- **Internal Team (n = 7):** This group comprised core members directly involved in the initial drafting of definitions and the coordination of the entire consensus process. They were leaders (YB, GF) and representatives of two large international consortia on vocal biomarkers: the Bridge2AI Voice Consortium and the eVoiceNet network. Their expertise spanned critical areas, including laryngology, speech-language pathology, digital epidemiology, AI, research methodology, and bioethics. This diverse composition ensured a strong foundational understanding of both clinical and technical aspects of vocal biomarkers.

- **Selected Group of Experts (n = 5):** This group consisted of international academic researchers with highly specialized expertise in speech analysis, clinical research, and the development and commercialization of vocal biomarkers. Their focused input provided in-depth technical and methodological critique, refining the precision and scientific accuracy of the definitions.

- **Broader Expert Group (n = 18):** This extended panel comprised a diverse array of researchers, clinicians, voice AI researchers, and stakeholders from a wide range of relevant disciplines (Clinical and Medical Sciences, Speech and Language Sciences, Engineering and Technical Fields, Statistics and Methodology, Regulation, Ethics, and

Translation). This group provided a comprehensive review and validation of the definitions from various practical and theoretical standpoints, ensuring their applicability across the broader scientific and clinical landscape.

Iterative Process

The process was organized into five iterative rounds of review and feedback, systematically designed to progressively refine the definitions and achieve strong consensus among all participants (see Figure 1; Supp Table 1).

- **Initial Draft and First Round of Review:** The internal team initiated the process by drafting preliminary definitions of vocal biomarkers. A comprehensive review informed this of existing literature and terminology. Following this, team members individually reviewed and provided comments on the draft. A subsequent group meeting was held to discuss and consolidate all feedback, leading to the first set of revisions (Revision #1).

- **Second Round of Review:** The revised definitions (Revision #1) were then shared with both the internal team and the selected group of experts. Participants provided written feedback, which was then discussed in an online workshop. This workshop focused on identifying points of agreement and divergence, and the collective input led to a second set of revisions (Revision #2).

- **Third Round of Review:** The updated definitions (Revision #2) were again reviewed by the internal team and the selected experts. A second online workshop was convened to refine terminology further and address any remaining ambiguities or concerns. This meticulous discussion resulted in a third revision (Revision #3), which brought the

definitions closer to a final consensus.

- **Fourth Round of Review:** The extensively revised definitions (Revision #3) were presented to the broader expert group for review and finalization. An in-person workshop was conducted at the Voice AI Conference in Tampa, April 2025, during which participants formally voted by a show of hands on their agreement or disagreement with each presented definition. Any definition receiving disagreement from 25% or more of the participants was subjected to further discussion and revision. This iterative refinement continued until a consensus was achieved for all definitions.

- **Fifth Round of Review and Finalization**: Feedback from the workshop was further integrated into a final set of definitions developed by the internal team. Following this final round, all co-authors reviewed and approved the definitions and the overarching framework presented in this paper.

This rigorous process ensured that all concerns were thoroughly addressed through collective discussion before the definitions were finalized.

Results

To capture the layered complexity of vocal biomarkers while maintaining clarity and usability, the definitions were structured across different levels of granularity, forming a continuum model. This hierarchical approach enables both a broad conceptual understanding and precise categorization based on the underlying physiological and cognitive systems involved in voice production.

- **Level 0** represents broad, overarching terms in the biomarker domain. These foundational definitions are intended to frame the scope of discussion and provide shared reference points across diverse disciplines and technologies.

- **Levels 1 to 4** provide more specific and structured definitions of terms that fall under the umbrella of vocal biomarkers. These are categorized based on the stage or domain of vocal production and analysis, ranging from the earliest physiological processes to higher-level cognitive functions (see Figure 2 and Figure 3).

It is crucial to understand that these categories are not mutually exclusive and often overlap, reflecting the interconnected nature of human voice production. Furthermore, extracted features can be interpreted as different types of biomarkers, depending on the specific context in which they are applied.

Level 0 Definitions

Table 1 establishes the foundational terminology for the entire framework, ensuring that readers from diverse backgrounds, including clinicians, engineers, researchers, and regulators, have a

common understanding of "Biomarker," "Digital Biomarker," and "Vocal Biomarker" before delving into more specific categories (see Supp Figure S1). This clarity is paramount for interdisciplinary discourse and regulatory alignment.

Level 1-4 Definitions

Table 2 breaks down the complex domain of vocal biomarkers into actionable, physiologically grounded categories. By providing clear definitions, examples, and illustrations, it serves as a practical guide for researchers and clinicians to accurately classify and interpret vocal features, thereby reducing ambiguity and facilitating standardized research design and reporting.

Discussion

This study presents the first stage of the VOCAL (Vocal Biomarker Guidelines for Ontology, Classification, Application, and Logistics) initiative, offering foundational definitions that aim to address longstanding terminological fragmentation in the field of vocal biomarkers. By introducing a layered, physiologically grounded vocabulary organized across five hierarchical levels, the framework establishes a shared reference system that facilitates interdisciplinary collaboration, supports regulatory alignment, and promotes scientific rigor. In doing so, it fills a critical gap in the biomedical and digital health landscape, where vocal biomarkers are increasingly positioned as scalable, non-invasive tools for the early detection and monitoring of diverse health conditions.

Rethinking Voice Data and Analytics Through a Unified Lexicon

This study's most consequential contribution is the reframing of "vocal biomarkers" not as a monolithic or interchangeable term, but as a continuum of overlapping physiological and cognitive signals. The framework intentionally avoids drawing rigid boundaries, opting instead for a layered structure that mirrors the interconnected nature of voice production, encompassing breath and muscle control, as well as language, memory, and emotion. This continuum model provides researchers and clinicians with a shared yet flexible vocabulary that can accommodate both scientific precision and real-world variability.

What makes this framework novel is its capacity to capture complexity without collapsing it. The five-level structure, from Level 0's foundational definitions (biomarker, digital biomarker, vocal biomarker) to Level 4's cognitive and language biomarkers, helps differentiate vocal signals not only by signal type, but also by the physiological or cognitive systems they reflect. For instance, Level 1 (cardio-respiratory acoustic biomarkers) isolates signals, such as breath pauses or

cough acoustics, which are distinct from Level 2's voice-source features, including pitch and harmonicity. Level 3 focuses on articulatory dynamics (tongue, lip, and jaw coordination), while Level 4 examines how language and social cognition are conveyed through lexical and paralinguistic cues. Importantly, this stratification does not imply mutually exclusive categories. The framework recognizes that many vocal features traverse multiple domains. Pitch, for example, may simultaneously carry phonatory, prosodic, and emotional information depending on the context. By allowing such overlaps, the model resists the oversimplification that often characterizes both academic studies and commercial applications in the voice-AI space.

This layered approach also acts as a conceptual correction to the field's current fragmentation. Too often, research on "vocal biomarkers" has used the term without clarifying what part of the vocal system is being measured, or for what purpose. This imprecision limits interoperability between datasets, obscures reproducibility, and undermines both clinical translation and regulatory review. The proposed definitions address this gap by creating clear referents across disciplines while still allowing interpretive flexibility. Beyond classification, the framework has normative and regulatory implications. By standardizing language across a fast-evolving and commercially saturated space, it challenges the opacity of proprietary systems that often treat voice as an undifferentiated input stream. A common lexicon, co-produced by experts across laryngology, AI, neuroscience, ethics, and clinical research, fosters transparency and makes space for critical scrutiny of what is being measured, how, and why. In this way, the framework functions not only as a taxonomic scaffold but as an epistemic and ethical intervention. It realigns the field around grounded, interpretable, and shared understandings of voice and its relationship to health, an essential foundation as vocal biomarkers move from exploratory research toward clinical integration.

Limitations

While this framework represents a significant step toward definitional coherence in vocal biomarker research, several limitations must be acknowledged. First, the process, although interdisciplinary and international, remains in its early stages. It centered primarily on researchers and professionals already embedded in consortia or academic networks. The voices of patients, regulators, and public health decision-makers from underrepresented regions were missing. Their perspectives are essential, particularly when voice data intersects with issues of trust, access, and cultural specificity. Second, the framework is intentionally definitional; it does not yet establish guidance on study design, feature extraction protocols, or validation metrics. While the definitions offer much-needed clarity on what is being measured, they do not resolve how those measurements should be standardized, interpreted, or integrated into clinical workflows. Questions regarding longitudinal stability, generalizability across languages and populations, and device dependence of vocal features remain outside the scope of this initial effort. Third, the flexible continuum approach (while a strength conceptually) may present practical challenges in operationalization. Some features can be categorized at more than one level, depending on their context. This ambiguity is intentional and reflects the non-discrete nature of vocal production, but it may complicate automated classification, regulatory labeling, or systematic reviews that require sharper boundaries. Finally, while the consensus process was rigorous and iterative, it does not constitute empirical validation. Whether these definitions improve reproducibility, regulatory acceptance, or interdisciplinary collaboration remains to be seen. The framework should thus be seen as provisional and living: a starting point for dialogue, not a final verdict.

Contribution to Vocal Biomarkers Development and Clinical Uptake

The definitional framework proposed in this paper addresses a longstanding barrier in the field of vocal biomarkers: the lack of precise and consistent terminology. By structuring vocal

biomarkers along a physiologically grounded continuum, the framework offers a common language for describing and interpreting voice-based health data. This contribution extends beyond academic precision, laying the groundwork for broader adoption, accountability, and clinical relevance.

Standardization acts as a catalyst for trust and uptake across the healthcare ecosystem. When definitions are unambiguous and widely accepted, clinicians are better equipped to interpret findings, patients are more likely to trust voice-based technologies, and regulators can more efficiently evaluate and approve them. This creates a reinforcing cycle: clarity builds trust, trust encourages adoption, and adoption drives further research, innovation, and investment. The framework also addresses a significant challenge in digital health: the opacity of proprietary systems. Open, transparent definitions reduce the "black box" problem by making it easier to understand what is being measured, how it is measured, and why. This fosters reproducibility, supports ethical governance, and enables cross-study comparisons, thereby reducing duplication and accelerating progress. For patients and clinicians alike, shared terminology enhances communication and facilitates the meaningful use of healthcare information. For third-party payers (e.g., private and public insurers) and regulators, standardization reduces uncertainty and facilitates alignment with existing biomarker approval pathways. In all cases, the framework advances the conditions necessary for responsible innovation and sustainable integration of vocal biomarkers into healthcare systems.

Future Research

This framework offers a foundation, not an endpoint. Future research should test the practical utility of the continuum model across various use cases, examining whether it improves consistency in feature labeling, interpretability, and reproducibility in both research and clinical settings. Applying the definitions in comparative studies may clarify their impact on model

transparency and cross-study compatibility. Methodological extensions are also needed, which include reporting standards aligned with the Levels, protocols for labeling multi-level features, and guidance on handling ambiguous cases. Embedding the framework into shared methodological infrastructures (such as master protocols) would support broader uptake and interoperability. Cultural, linguistic, and ethical dimensions must also be taken into account. Vocal features are shaped by more than physiology; they carry social and contextual meaning. Research must examine how definitions perform across diverse populations and communication norms, and avoid encoding bias into measurement or interpretation. Participatory approaches will be crucial in ensuring that voice-based tools are designed and deployed fairly and equitably. Finally, alignment with regulatory standards is essential. Mapping the framework to existing evidentiary categories (such as digital endpoints or AI-based diagnostics) can help establish vocal biomarkers as clinically valid and ethically deployable tools, not speculative proxies. As the field matures, this definitional groundwork should evolve alongside it, anchoring technical progress in clarity, accountability, and shared understanding.

Conclusion

This paper presents the first structured, consensus-based definitional framework for vocal biomarkers, responding to a critical need for clarity, consistency, and interoperability in a rapidly evolving field. Developed through a rigorous, multi-stage process involving international experts across disciplines, the framework offers a shared vocabulary that reflects the physiological and cognitive dimensions of voice production. By anchoring vocal biomarkers in a layered, interpretable structure, this work provides a foundation for interdisciplinary collaboration, ethical development, and clinical translation. It helps unify fragmented research efforts, supports reproducibility, and facilitates regulatory alignment, paving the way for future guidelines, validation standards, and implementation practices. As voice continues to emerge as a rich

source of health-relevant data, a common language is indispensable. This framework offers more than definitional clarity; it offers a path toward trustworthy, patient-centered, and clinically impactful voice-based health technologies.

Acknowledgments

We thank Claire Bortolotto for her contributions during selected meetings of the first round of review.

References

1. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. Digit Biomark. 2021;5(1):78-88. doi:10.1159/000515346

2. Bensoussan Y, Elemento O, Rameau A. Voice as an AI Biomarker of Health—Introducing Audiomics. JAMA Otolaryngol Neck Surg. 2024;150(4):283-284. doi:10.1001/jamaoto.2023.4807

3. Fagherazzi G, Bensoussan Y. The Imperative of Voice Data Collection in Clinical Trials. Digit Biomark. 2024;8(1):207-209. doi:10.1159/000541456

4. Kalia A, Boyer M, Fagherazzi G, Bélisle-Pipon JC, Bensoussan Y. Master protocols in vocal biomarker development to reduce variability and advance clinical precision: a narrative review. Front Digit Health. 2025;7:1619183. doi:10.3389/fdgth.2025.1619183

5. Patel R, Price N, Bahr R, et al. Summary of Keynote Speeches from the 2024 Voice AI Symposium, presented by the Bridge2AI-Voice Consortium. Front Digit Health. 2024;6. doi:10.3389/fdgth.2024.1484503

6. Hajjar I, Okafor M, Choi JD, et al. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. Alzheimers Dement Amst Neth. 2023;15(1):e12393. doi:10.1002/dad2.12393

7. Bauser M, Kraus F, Koehler F, et al. Voice Assessment and Vocal Biomarkers in Heart Failure: A Systematic Review. Circ Heart Fail. Published online April 24, 2025:e012303. doi:10.1161/CIRCHEARTFAILURE.124.012303

8. Ayadi H, Elbéji A, Despotovic V, Fagherazzi G. Digital Vocal Biomarker of Smoking Status Using Ecological Audio Recordings: Results from the Colive Voice Study. Digit Biomark. 2024;8(1):159-170. doi:10.1159/000540327

9. Elbeji A, Aguayo G, Fischer A, Fagherazzi G. Identification of vocal biomarkers for screening diabetes and monitoring health of people with diabetes: preliminary results from the COLIVE voice study. Diabetes Technol Ther. 2022;24(Suppl.1):A224-A224. doi:10.1089/dia.2022.2525.abstracts

10. Kiyoshige E, Ogata S, Kwon N, et al. Developing and testing AI-based voice biomarker models to detect cognitive impairment among community dwelling adults: a cross-sectional study in Japan. Lancet Reg Health West Pac. 2025;59:101598. doi:10.1016/j.lanwpc.2025.101598

11. Mazur A, Costantino H, Tom P, Wilson MP, Thompson RG. Evaluation of an AI-Based Voice Biomarker Tool to Detect Signals Consistent With Moderate to Severe Depression. Ann Fam Med. 2025;23(1):60-65. doi:10.1370/afm.240091

12. Watase T, Omiya Y, Tokuno S. Severity Classification Using Dynamic Time Warping-

Based Voice Biomarkers for Patients With COVID-19: Feasibility Cross-Sectional Study. JMIR Biomed Eng. 2023;8:e50924. doi:10.2196/50924

13. Kerwagen F, Bauser M, Baur M, et al. Vocal biomarkers in heart failure-design, rationale and baseline characteristics of the AHF-Voice study. Front Digit Health. 2025;7:1548600. doi:10.3389/fdgth.2025.1548600

14. Lin YC, Yan HT, Lin CH, Chang HH. Identifying and Estimating Frailty Phenotypes by Vocal Biomarkers: Cross-Sectional Study. J Med Internet Res. 2024;26:e58466. doi:10.2196/58466

15. Roland V, Huet K, Harmegnies B, Piccaluga M, Verhaegen C, Delvaux V. Vowel production: a potential speech biomarker for early detection of dysarthria in Parkinson's disease. Front Psychol. 2023;14:1129830. doi:10.3389/fpsyg.2023.1129830

16. Gaur S, Kalani P, Mohan M. Harmonic-to-noise ratio as speech biomarker for fatigue: K-nearest neighbour machine learning algorithm. Med J Armed Forces India. 2024;80(Suppl 1):S120-S126. doi:10.1016/j.mjafi.2022.12.001

17. Sterner B. Explaining ambiguity in scientific language. Synthese. 2022;200(5):354. doi:10.1007/s11229-022-03792-x

18. Livieri G, Mangina E, Protopapadakis ED, Panayiotou AG. The gaps and challenges in digital health technology use as perceived by patients: a scoping review and narrative meta-synthesis. Front Digit Health. 2025;7. doi:10.3389/fdgth.2025.1474956

19. Key Factors Behind IT Implementation Failures in Mid-Sized Healthcare Settings - CapMinds. May 5, 2025. https://www.capminds.com/blog/key-factors-behind-it-implementation-failures-in-mid-sized-healthcare-settings/

20. Babrak LM, Menetski J, Rebhan M, et al. Traditional and Digital Biomarkers: Two Worlds Apart? Digit Biomark. 2019;3(2):92-102. doi:10.1159/000502000

21. Schmidt J, Schutte NM, Buttigieg S, et al. Mapping the regulatory landscape for artificial intelligence in health within the European Union. Npj Digit Med. 2024;7(1):229. doi:10.1038/s41746-024-01221-6

22. Bélisle-Pipon JC, Powell M, English R, et al. Stakeholder perspectives on ethical and trustworthy voice AI in health care. Digit Health. 2024;10:20552076241260407. doi:10.1177/20552076241260407

23. Tan RKJ, Wu D, Day S, et al. Digital approaches to enhancing community engagement in clinical trials. NPJ Digit Med. 2022;5(1):37. doi:10.1038/s41746-022-00581-1

Figures legends

Figure 1. Flowchart of the iterative process to establish the definitions of key terms of the VOCAL initiative

The process included an initial internal team draft and first review, two rounds of review with selected experts, a consensus workshop with a broader expert group, and a final review and approval. The number of participants at each stage is indicated (n).

Figure 2. Hierarchical continuum of vocal biomarkers: from voice production to high-level cognition

The pyramid illustrates the hierarchy of vocal biomarkers, organized by physiological drivers and complexity. **Level 1** biomarkers (blue) represent cardio-respiratory acoustic biomarkers driven by respiration and aerodynamics. **Level 2** biomarkers (green) reflect voice biomarkers arising from phonation. **Level 3** biomarkers (yellow) capture speech and articulatory biomarkers shaped by articulation and resonance. **Level 4** biomarkers (orange) represent cognitive and language biomarkers arising from high-order neurological processes. Complexity increases from the base to the apex of the pyramid. Corresponding anatomical regions associated with each biomarker type are shown on the left. Created in https://BioRender.com

Figure 3. Vocal Biomarkers Levels 0-4

Vocal biomarkers can be categorized across four levels based on their physiological and cognitive origins. **Level 0: Biomarker, Digital Biomarker, Vocal biomarker** represent broad, overarching terms. **Level 1: Cardio-Respiratory Acoustic Biomarkers (CRA)** relate to

acoustic signals associated with respiration and cardiac function (e.g., breathing sounds, cough acoustics). **Level 2: Voice Biomarkers** pertain to features of phonation and vocal fold vibration (e.g., harmonic-to-noise ratio, cepstral peak prominence, fundamental frequency). **Level 3: Speech/Articulatory Biomarkers** involve articulatory processes and speech motor control (e.g., articulation rate, speech rate, articulatory precision, speech prosody and rhythm). **Level 4: Cognitive/Language Biomarkers** reflect the content of speech, including lexical access, word pairing, syntax, and paralinguistic elements such as emotion, cultural cues, and dialects (e.g., vocabulary, syntax). Created in https://BioRender.com

Tables

Table 1. Level 0 Definitions

Term	Definition	Explanatory Notes/Examples
Biomarker	A measurable biological substance or characteristic whose detection or evaluation provides information about the function or state of a system or organ in the body. It can be used to understand normal or altered biological processes, and may serve roles in risk prediction, screening, diagnosis, monitoring, treatment response measurement, or prognosis.	There are various types of biomarkers, including physiological, serological, imaging, and genetic ones. For instance, HbA1c serves as a blood biomarker for diabetes. Another example of a biomarker is the presence of the BRCA gene, which indicates a genetic predisposition to certain types of breast cancer. In line with regulatory definitions, a biomarker is formally recognized only after undergoing appropriate validation (analytical and clinical). However, within the research and innovation context, especially in emerging areas such as voice-based health technologies, the term "biomarker" is commonly used pre-validation to denote a candidate indicator under investigation.
Digital Biomarker	A biomarker derived from data collected using a digital device such as a wearable, portable, or implantable device.	For example, a heart rate measured from a smartwatch is a digital biomarker, and an ECG measured by a wearable is a digital biomarker. A digital biomarker is a digital measure that can be used as a digital endpoint in a clinical study. However, not all digital measures are digital biomarkers.
Vocal Biomarker (overarching term, includes voice, speech, linguistic, paralinguistic, and prosody)	A vocal biomarker is an overarching term for defining biomarkers derived from audio recordings of the voice, speech, linguistic, and paralinguistic	It is important to understand that these should not represent mutually exclusive categories, as they can overlap and should be interpreted more as a continuum. Moreover,

Term	Definition	Explanatory Notes/Examples
	systems. These are collected through acoustic data recording and can be analyzed using traditional acoustic analysis or digital signal processing and AI methods. This overarching term includes biomarkers that are impacted by the process of creating speech from our lungs and cardio-respiratory system to our larynx, resonators, and brain, which helps with lexical access and influences the content of our speech.	extracted features can be interpreted as different types of biomarkers, depending on the context. Example: F0 can be used as a voice biomarker in dysphonia or voice disorders, and can also be considered in mental health screening for depression.

Table 2. Level 1-4 Definitions

Term	Definition	Examples	Illustrations
Cardio-Respiratory Acoustic (CRA) biomarkers (Level 1)	Cardio-respiratory acoustic (CRA) biomarkers relate to acoustic signals pertaining to respiration and cardiac function. They represent the earliest stage in the speech production pathway and reflect physiological processes that influence vocalization.	Breathing sounds, cough acoustics, and pauses caused by breath support.	A cough with low intensity may reflect reduced respiratory drive or weakened supportive muscles. Acoustic analysis of coughs or breathing may indicate respiratory illness or muscular disorders. Frequent pauses in speech may reflect dyspnea or reduced breath control.
Voice biomarkers (Level 2)	Voice biomarkers relate to features from the source of phonation: the lungs pushing air through the vocal folds, which vibrate to generate sound, later shaped by the vocal	Harmonic-to-noise ratio (HNR), cepstral peak prominence (CPP), fundamental frequency (F0).	In laryngitis, vocal fold inflammation can cause a rough or hoarse voice quality, measurable using CPP. Parkinson’s disease can affect prosody, leading to monotonous speech with reduced pitch variability (F0

Term	Definition	Examples	Illustrations
	tract (throat, mouth, nose).		range) and inflection.
Speech/Articulatory biomarkers (Level 3)	Speech/Articulatory biomarkers relate to articulatory processes and speech motor control. They reflect the coordination and functioning of articulators (tongue, lips, soft palate, etc.), which are regulated by neurological control.	Articulation rate, speech rate, syllable precision, consonant clarity.	In ALS, tongue weakness may impair the articulation of syllables like “Ka” or “Ga.” Variations in rhythm and timing can indicate altered motor coordination or speech effort.
Cognitive/Language biomarkers (Level 4)	Cognitive/Language biomarkers relate to the content and meaning of speech, reflecting higher-level cognitive functions, language processing, and emotional or social communication cues.	***Linguistic Biomarkers*** Refer to the content of our speech, our lexical access, the words we choose in a sentence, and how we pair words together to form sentences. They can also include syntax, grammar, semantics, and vocabulary concepts. ***Paralinguistic Biomarkers*** include non-verbal communication elements, such as conveying emotion, cultural cues, and dialects.	**Linguistic Illustration:** Reduced vocabulary diversity in Alzheimer’s disease or syntactic simplification in aphasia. **Paralinguistic Illustration:** Intonation patterns signaling questions or emotional states (e.g., rising pitch) may reflect cognitive or affective modulation in conditions like autism spectrum disorder.

Initial Draft and First Round of Review

Internal team (n = 7)

- Internal team drafted preliminary definitions (literature review-informed)
- Individual reviews and comments by internal team
- Internal team meeting to consolidate feedback
- First set of revisions (Revision #1)

Second Round of Review

Selected experts (n = 5)

- Revision #1 shared with internal team and selected experts
- Written feedback collected
- Online workshop to discuss feedback, align on terminology
- Second set of revisions (Revisions #2)

Third Round of Review

Selected experts (n = 5)

- Revision #2 reviewed by internal team and selected experts
- Second online workshop held to refine definitions
- Third set of revisions (Revision #3)

Forth Round of Review (Consensus Workshop)

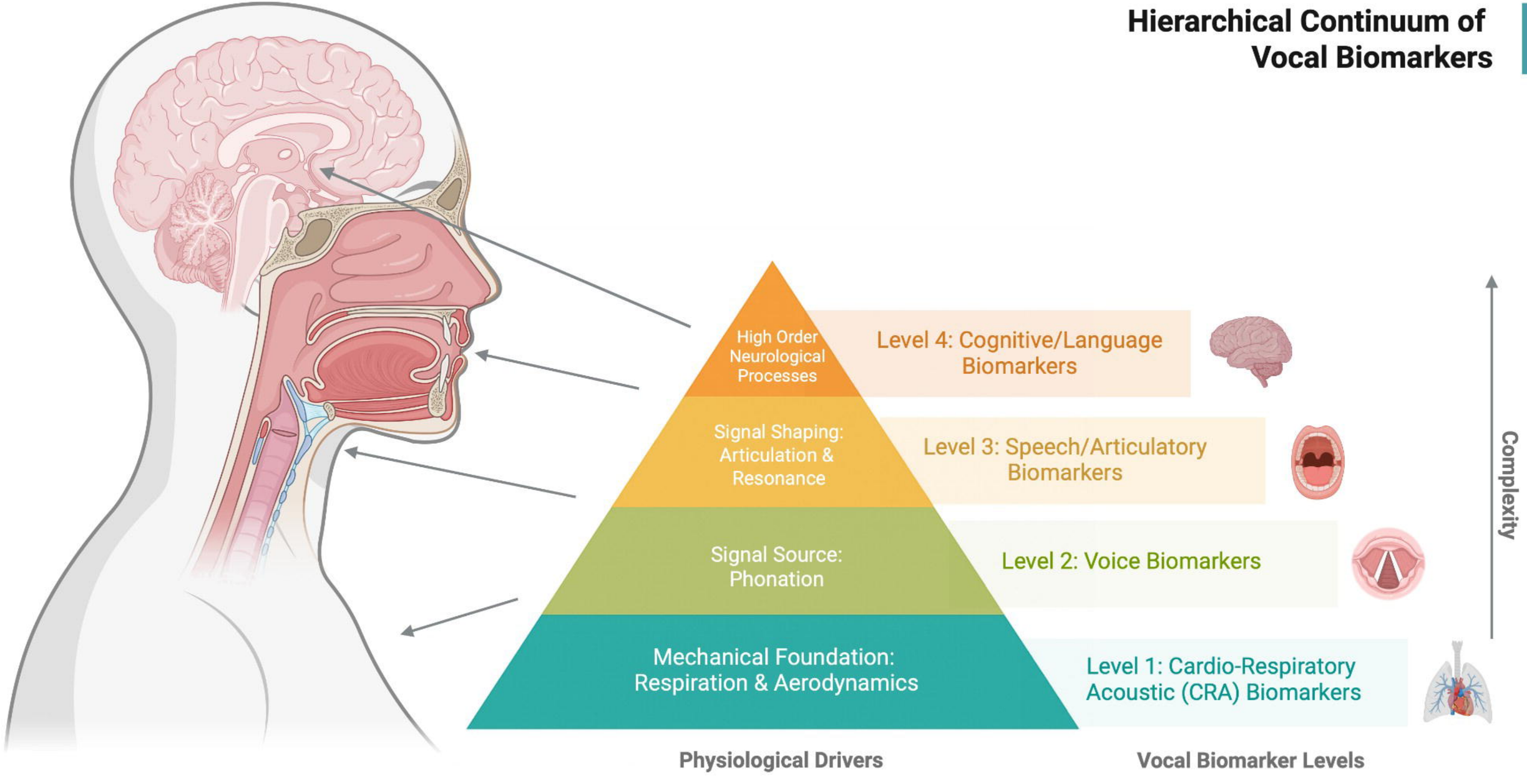
Broader expert group (n = 18)

- Revision #3 presented to broader expert group
- In-person workshop at Voice AI Conference, Tampa (April 2025)
- Formal voting on each definition
- Definitions with ≥25% disagreement revised further
- Iterative discussion until full consensus achieved

Fifth Round of Review and Finalization

- Final feedback integrated by internal team
- All co-authors reviewed and approved definitions and framework

Hierarchical Continuum of Vocal Biomarkers



VOCAL Initiative Vocal Biomarker Definitions Levels 0-4

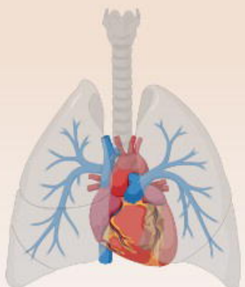
Level 0 Definitions

Biomarker: A measurable biological substance or characteristic whose detection or evaluation provides information about the function or state of a system or organ in the body.

Digital Biomarker: A biomarker derived from data collected using a digital device such as a wearable, portable, or implantable device.

Vocal Biomarker: An overarching term for defining biomarkers derived from audio recordings of the voice, speech, linguistic, and paralinguistic systems

Level 1: Cardio-Respiratory Acoustic Biomarkers (CRA)



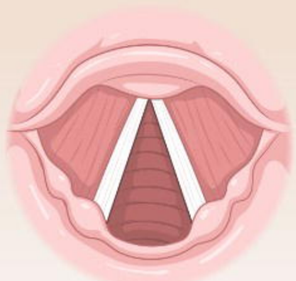
Definition

- Relates to acoustic signals pertaining to respiration and cardiac function.

Examples

- Breathing sounds
- Cough acoustics

Level 2: Voice Biomarkers



Definition

- Relate to features of the source of phonation and vocal fold vibration

Examples

- Harmonic-to-noise ratio (HNR)
- Cepstral Peak Prominence (CPP)
- Fundamental Frequency (F0)

Level 3: Speech/ Articulatory Biomarkers

Definition

- Relate to articulatory processes and speech motor control

Examples

- Articulation rate
- Speech rate
- Articulatory Precision
- Speech prosody and rhythm

Level 4: Cognitive/ Language Biomarkers

Definition

- Relates to the content of speech, lexical access, the words that are chosen and how words are paired to form sentences.

Examples

- Vocabulary
- Syntax