

VIEWPOINT

Yaël Bensoussan, MD, MSc

USF Health Voice Center, Department of Otolaryngology–Head & Neck Surgery, University of South Florida Health Morsani College of Medicine, Tampa.

Olivier Elemento, PhD

Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, New York.

Anaïs Rameau, MD, MPhil

Sean Parker Institute for the Voice, Department of Otolaryngology–Head and Neck Surgery, Weill Cornell Medicine, New York, New York.



Multimedia

Corresponding

Author: Yaël Bensoussan, MD, MSc, Division of Laryngology, USF Health Voice Center, Department of Otolaryngology–Head & Neck Surgery, University of South Florida Morsani College of Medicine, 13330 USF Laurel Dr, Tampa, FL 33609 (yaelbensoussan@usf.edu).

Voice as an AI Biomarker of Health—Introducing Audiomics

Voice, speech, and respiratory sounds provide important clinical insights into patients' health status. In the age of artificial intelligence (AI), patients' audio recordings are being investigated as digital biomarkers for early detection of a broad range of conditions, including laryngeal pathology, neurological and psychological disorders, head and neck cancers, and diabetes. Besides neurologists, speech language pathologists, and interventionalists, otolaryngologists also have unique perspectives and expertise on voice, speech, and respiratory sounds that can fuel this line of research and innovation.

While the potential of using voice as a biomarker of health has been explored for decades, development of recent technology such as machine learning (ML) (a branch of AI focusing on predictive algorithms that learn from data without explicit instructions) allows for efficient analysis of a voice data and makes discovery of scalable acoustic biomarkers a possibility. The impact of such a discovery on patient care could be substantial, including in screening, diagnosis, remote monitoring, and development of new digital end points for clinical trials.

Although the future of voice biomarkers is promising, there remain important limitations to broad integration into clinical care. Within academic research, many studies^{1,2} remain at the level of proof of concept with small- to medium-sized datasets often using voice as the only data type. Comparing studies and pooling data are challenging tasks due to the lack of standards in how we collect voice and speech data. Within the industry, datasets are often private, limiting the ability to audit ML models and the accuracy of training data labels.

Herein, we highlight the need to define and create standards for “audiomics,” the interdisciplinary field of audio analysis applied to biomedicine to identify unique audio biomarkers of health and disease. Recent advances in data science have set the stage for the emergence of new “omics,” that is, data science subfields founded on a large amount of data representing the structure or function of a biological system.³ Examples of such omics subfields include radiomics, genomics, proteomics, and, more recently, videomics, referring to the application of omics in video analysis, such as endoscopy.⁴ Videomics was developed for gastrointestinal endoscopy for lesion identification during colonoscopy and has recently been expanded to endoscopy videos in otolaryngology.⁵ Central to omics are efforts toward standardization of data collection, storage, and analysis and ongoing discussion of ethical and regulatory challenges, which will take simple acoustic analysis to a new level. Otolaryngologists have a rare opportunity with audiomics to contribute to cutting-edge medicine by sharing how voice, speech, and respiratory sounds may be interpreted in the context of human health and by guiding the collection of acoustic data

at scale to offer populationwide solutions. Tangible use cases in otolaryngology include voice screening for detection and monitoring of laryngeal cancer in at-risk populations for timely referral, or remote voice monitoring in spasmodic dysphonia for assessing botulinum toxin response to fine-tune and personalize treatment.

Among digital biomarkers, voice, speech, and respiratory sounds are particularly attractive due to their noninvasive, accessible, and low-cost collection in the setting of recording capability via computers or smartphones.¹ These digital data have come in focus in the age of automated speech recognition systems and large language models, with unparalleled ability to analyze voice and speech. Furthermore, digital biomarkers are increasingly sought as clinical data outcome measures for therapeutics, particularly in the pharmaceutical industry. Such data could overcome limitations of clinical trials and research ethics concerns, such as time required for participant enrollment, challenges related to intervention delivery, data collection burden, and lack of diversity among enrolled participants.

Many industry giants, such as Pfizer, Mozilla, Amazon, Google, and Apple, and start-ups are investing millions in research and development of voice, speech, and respiratory sound biomarkers.⁶ Although the industry has access to vast amounts of voice data from users, it lacks access to accurate clinical data or demographic information. This deficiency limits the ability to annotate data for AI training and to compare algorithmic outputs with criterion standard medical assessment and prevents external auditing of datasets used for training to ensure fair representation of diverse populations.

Despite significant advances, important limitations currently prevent implementation of voice biomarkers in clinical care. Despite considerable literature^{1,2} published especially in the past decade, there is a lack of prospective study and validation of AI algorithms in audiomics on external datasets. These factors, compounded by the limited size, quality, and diversity of existing open-source datasets, likely explain the lack of validated and US Food and Drug Administration–approved AI algorithms for disease detection and monitoring in this space. When ML applications from screening and diagnosis of laryngeal cancer were considered, researchers found that most studies reported training datasets of fewer than 300 patients, and only 2 studies used more than 1 data modality to train their models, limiting generalization of the results.⁷ Furthermore, audiomics will inevitably need to be integrated in multi-omics efforts, in which audio data are analyzed together with other health data, such as clinical, imaging, and even genomic data, for improved accuracy. When a patient presents with a dysphonic voice, we clinicians inquire about duration of symptoms and smoking history to gauge risk of laryn-

geal cancer. Machine learning models trained on voice data need the same type of multimodal integration to achieve scalable accuracy.

For audiomics to be implemented and reach its full public health potential, the following tripartite agenda must be met: (1) standards for audiomics data acquisition, interoperability, and AI readiness must be defined; (2) diverse teams with broad skills and expertise, including clinicians, engineers, bioethicists, and social scientists, must collaborate in this complex field; and (3) the unique ethical and legal challenges in audiomics linked to the Health Insurance Portability and Accountability Act, potential reidentification of patients by their voice, voice hacking, and data ownership call for a new governance framework for issues concerning diversity, data handling, and privacy safeguards.

Through the Bridge2AI program, an endeavor to leverage team science in AI sponsored by the National Institutes of Health Common Fund, our team, Bridge2AI-Voice, gathered 50 multidisciplinary experts from 12 North American institutions to generate high-quality, ethically sourced datasets for biomedical research to advance the field of audiomics. Besides our primary deliverable to build a publicly available database of 30 000 human voices, our

team aims to address current gaps in the burgeoning field of audiomics by (1) ensuring quality and accuracy of acoustic data and associated clinical data through clinical validation and evidence-based recording protocols; (2) aiming for interoperability of data by creating new bioinformatics standards for voice data; (3) creating benchmarks of diversity metrics and minimizing risk of algorithmic bias; (4) defining ethical and legal norms safeguarding patients' data while supporting data-sharing efforts; (5) developing infrastructure for audiomics data storage sharing within and across research institutions; and (6) formulating training pathways for scientists, engineers, bioethicists, and clinicians to develop skills in audiomics.

The only scalable way to achieve such a complex endeavor and implement change beyond the current hype around acoustic biomarkers is to bring experts from different fields, including otolaryngologists, speech pathologists, data scientists, AI engineers, bioethicists, and acousticians, to collaborate and build a solid, ethically sourced infrastructure for voice data collection linked to other health data. In this way, the Bridge2AI-Voice data generation project serves as a model from which current and future generations of audiomics researchers can learn.

ARTICLE INFORMATION

Published Online: February 22, 2024.
doi:10.1001/jamaoto.2023.4807

Conflict of Interest Disclosures: Dr Elemento reported owning equity in Owkin and in Volastra Therapeutics outside the submitted work. Dr Rameau reported receiving grants from the National Institute on Aging and owning equity in Perceptron Health, Inc, and Savorease, Inc, outside the submitted work. No other disclosures were reported.

Funding/Support: All authors are funded by Bridge2AI award OT2 OD032720 from the National Institutes of Health Common Fund.

Role of the Funder/Sponsor: The National Institutes of Health had no role in the preparation, review, or approval of the manuscript or the decision to submit the manuscript for publication.

Additional Contributions: We thank the following members of the Bridge2AI-Voice Collaborators for their contributions to the Bridge2AI program: Drs Bensoussan, Elemento, and Rameau as well as Jean-Christophe Bélisle-Pipon, PhD, Faculty of Health Science, Department of Health Ethics, Simon Fraser University; David A. Dorr, MD, MS,

Department of Health Informatics and Clinical Epidemiology, Oregon Health Sciences University; Satrajit S. Ghosh, PhD, McGovern Institute, Massachusetts Institute of Technology; Alistair Johnson, PhD, Division of Biostatistics, Hospital for Sick Children; Philip R. O. Payne, PhD, Institute for Informatics, Data Science and Biostatistics, Washington University School of Medicine in St Louis; Maria Powell, PhD, CCC-SLP, Department of Otolaryngology-Head & Neck Surgery, Vanderbilt University Medical Center; Vardit Ravitsky, PhD, The Hastings Center; and Alexandros Sigras, MS, Englander Institute for Precision Medicine, Weill Cornell Medicine. None were directly compensated for their contributions.

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. Idrisoglu A, Dallora AL, Anderberg P, Berglund JS. Applied machine learning techniques to diagnose voice-affecting conditions and disorders: systematic literature review. *J Med Internet Res*. 2023;25:e46105. doi:10.2196/46105

3. Micheel CM, Nass SJ, Omenn GS, et al. Omics-based clinical discovery: science, technology, and applications. In: *Evolution of Translational Omics: Lessons Learned and the Path Forward*. National Academies Press; 2012. doi:10.17226/13297
4. Dai X, Shen L. Advances and trends in omics technology development. *Front Med (Lausanne)*. 2022;9. doi:10.3389/fmed.2022.911861
5. Paderno A, Holsinger FC, Piazza C. Videomics: bringing deep learning to diagnostic endoscopy. *Curr Opin Otolaryngol Head Neck Surg*. 2021;29(2):143-148. doi:10.1097/MOO.0000000000000697
6. Apple introduces new features for cognitive accessibility, along with Live Speech, Personal Voice, and Point and Speak in Magnifier. Apple Newsroom (Canada). May 16, 2023. Accessed July 24, 2023. <https://www.apple.com/ca/newsroom/2023/05/apple-previews-live-speech-personal-voice-and-more-new-accessibility-features/>
7. Bensoussan Y, Vanstrum EB, Johns MM III, Rameau A. Artificial intelligence and laryngeal cancer: from screening to prognosis: a state of the art review. *Otolaryngol Head Neck Surg*. 2023;168(3):319-329. doi:10.1177/0194599822110839