Interspeech 2024
1-5 September 2024, Kos, Greece

Developing Multi-Disorder Voice Protocols: A team science approach involving clinical expertise, bioethics, standards, and DEI.

Anais Rameau[1], Satrajit Ghosh[2], Alexandros Sigaras[3], Olivier Elemento[4], Jean-Christophe Belisle-Pipon[5,] Vardit Ravitsky[6], Maria Powell[7], Alistair Johnson[8], David Dorr[9], Philip Payne[10], Micah Boyer[11], Stephanie Watts[11], Ruth Bahr[12], Frank Rudzick[13], Jordan-Lerner-Ellis[14], Shaheen Awan[15], Don Bolser[16], Bridge2AI-Voice Consortium, Yael Bensoussan[11]

[1]Department of Otolaryngology-Head and Neck Surgery, Weill Cornell Medicine, New York, New York, USA

[2]McGovern Institute, Massachusetts Institute of Technology, Boston, Massachusetts, USA

[3]Institute for Computational Biomedicine, Weill Cornell Medicine, New York, New York, USA

[4] Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, New York, USA

[5] Faculty of Health Sciences, Simon-Fraser University, Vancouver, British Columbia, Canada

[6] Hasting Center, Garrison, New York, USA

[7]Department of Otolaryngology-Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[8]Division of Biostatistics, Hospital for Sick Children, Toronto, Ontario, Canada

[9]Department of Medical Informatics, School of Medicine, Oregon Health & Science University, Portland, Oregon, USA

[10] Institute for Informatics, Washington University in St Louis, St Louis, Missouri, USA

[11]Department of Otolaryngology- Head & Neck Surgery, University of South Florida, Tampa, Florida, USA. [12] Department of Communication Sciences & Disorders, University of South Florida, Tampa, Florida, USA. [13]Faculty of Computer Science, Dalhousie University, Calgary, Manitoba Canada

[14]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

[15]Department of Communication Sciences & Disorders, University of Central Florida, Orlando, Florida, USA

[16.] Department of Physiological Sciences, University of Florida, Gainesville, Florida, USA

anr2783@med.cornell.edu, satra@mit.edu, als2076@med.cornell.edu, ole2001@med.cornell.edu, jean-christophe_belisle-pipon@sfu.ca, vardit.ravitsky@umontreal.ca, maria.e.powell@vumc.org, alistair@glowyr.ca, dorrd@ohsu.edu, prpayne@wustl.edu, micahboyer@usf.edu, srandall1@usf.edu, rbahr@usf.edu, frank@spoclab.com, jordan.lerner-ellis@sinaihealth.ca, Shaheen.awan@ucf.edu, bolser@ufl.edu, yaelbensoussan@usf.edu

10.21437/Interspeech.2024-1926

Abstract

The world of voice biomarkers is rapidly evolving thanks to the use of artificial intelligence (AI) allowing large-scale analysis of voice, speech, and respiratory sound data. The Bridge2AI-Voice project aims to build a large-scale, ethically sourced, and diverse voice database of human voices linked to health information to help fuel Voice AI research, dubbed Audiomics. The current paper describes the development of protocols of data acquisition across 4 different adult cohorts of disease (voice, respiratory, neurodegenerative diseases, mood, and anxiety disorders) using a Team Science approach for broader adoption by the research community and feedback. Demographic Surveys, Confounders Assessments, Acoustic tasks, validated patient-reported outcome (PRO) questionnaires and clinician-validated diagnostic questions were grouped in a common PART A across all cohorts and individual PART B, with cohort-specific tasks.

Index Terms: standardized voice protocol, disorders, consent, team science

1. Introduction

1.1. Current limitations in voice AI research

There is ample evidence today that information derived from voice recordings can be used for diagnosis and disease progression or recovery across many disorders [1]–[7]. Given current advances in machine learning and AI technologies, models based on such data would be able to leverage such information to provide scalable and low-cost disease assessment and tracking capabilities in local and remote settings for many healthcare applications [8]. A major challenge in the development of robust voice and speech-based AI models for such applications is hampered by several challenges.

1.1.1. Lack of accessible, high-quality, and diverse voice data for biomedical applications

While there are several large datasets available for speech recognition research, the number of accessible and clinically relevant datasets with voice, speech, and associated clinical metadata is relatively small. Most clinical voice and speech datasets remain inaccessible for lack of appropriate consent and sensitivity related to identifiability. When available, these datasets are often generated without coordination in isolated labs, are associated with a specific disorder and a specific set of hypotheses, varied in their recording quality and settings, and distributed with limited provenance. Most importantly, either metadata about diversity is lacking or the data are not representative of different health-related dimensions of a diverse population.

1.1.2. Lack of standardized and reusable protocols for data collection

Many of the existing datasets have been created without significant consideration for multi-disorder applications. Even datasets generated for a disorder do not use consistent protocols for collecting demographic information, clinical validation, and acoustic tasks. Data corresponding to these datasets are often collected using laboratory-specific or ad hoc acoustic data collection approaches in varying external environments, and downstream users are left with the challenging task of separating protocol-related and disorder-related sources of variability.

1.1.3. Need for multi-disorder and multimodal voice data

Voice disorders are often directly linked to variations in acoustics of voice and speech recordings. However, as human generation of vocalic sound relies on the brain, the respiratory system, and the intrinsic biophysical properties of the vocal and nasal tract, the acoustic output is also affected by neurological and psychiatric disorders. Therefore, to understand acoustic and linguistic variations in voice and speech recordings that are unique to or shared across disorders, an effort is needed to assemble a multi-disorder dataset using standardized protocols and containing detailed metadata. Such a dataset will help establish biomarkers by connecting different pathophysiology and trajectories of disorders with corresponding acoustic and linguistic features.

Brigde2AI-Voice aimed to address these challenges and develop a standardized presentation and collection protocol to link voice data to other health biomarkers and generate a large- scale, multi-institutional, multimodal, AI-ready, voice database that is ethically sourced and representative of a diverse population. The primary objective of this paper is to present the standardized protocols for data acquisition as well as how they were developed through a team science approach, together with ethical evaluations and considerations for diversity, equity, and inclusion (DEI) incorporated at each stage of the process.

2. Methods

The protocols for data acquisition were developed in an iterative and interactive fashion over the course of 6 months by a multidisciplinary, multi-institutional Team Science. Protocols for four adult disease cohorts were developed: 1. Voice disorders 2. Respiratory disorders 3. Mood disorders 4. Neurological Disorders.

2.1. Stakeholders involved in Team Science

The "unnamed" consortium is composed of the following stakeholders who worked in intricate collaboration through the development of the protocols. Cohort leads were then identified as leaders in their respective fields of diseases studied.

- Clinicians with expertise in specific disease cohorts: This group included providers with expertise in the 4 disease cohorts including laryngologists and speech pathologists, pulmonologists, neurologists, and mood disorder experts.
- Researchers with expertise in specific disease cohorts: This group included academic researchers who had published previous work in voice biomarkers for their respective disease cohorts.
- Bioethicists: This group included bioethicists and lawyers with expertise in artificial intelligence and biotechnology.
- Diversity experts: This group included researchers with an academic focus on diversity, inclusion, and accessibility in the context of innovation
- Bioinformaticians: This group included software developers, data scientists, machine learning, and UX/UI experts.

2.2. Data types for the protocols

Experts were separated by cohorts (Neuro, Mood, Voice, Respiratory) to develop the initial protocols separately based on literature review and clinical expertise over the course of 6 weeks. Each cohort lead searched the literature extensively to find the current standards to inform protocol development. Team science was tasked to develop protocols including the following data types:

- Demographics: The group was asked to include common demographic data and include other demographics that could affect voice and speech (ex: weight, socio-economic status, literacy status, etc)
- Past medical history (Pmhx): The group was asked to include common disorders with care being taken to include diseases and conditions that are known to affect voice and speech (ex: COPD, chronic sinusitis,

etc)

- Confounders: The group was asked to include confounders and social habits that are known to affect voice and speech (ex: Smoking status, hydration status, etc)
- Acoustic tasks: The group was asked to include common acoustic tasks performed for screening or diagnosis of the conditions studied in the clinical setting or research setting.
- Validated questionnaires and PROs: Patient-reported outcomes (PROs) and validated patient questionnaires commonly used in clinical or research practice with evidenced-based correlation with the diseases studied (ex: GAD-7 for anxiety or VHI-10 for dysphonia)
- Clinical Validation: the group was asked to develop a section including questions that would confirm the diagnosis and treatment obtained by a clinician
- "Gold Standards": The group was asked to add data modality that are used for confirmation or included in the basic work-up of the diseases studied (ex: MRI of the brain for Alzheimer's disease, pathology report for laryngeal cancer, pulmonary function test for Asthma, etc)

Table 1: Multi-Omics "gold standard" data

	Resp.	Voice	Neurological
Images	CT Chest	Laryngo-scopy	Brain MRI
Genomics			Whole Genome Sequencing
Pathology		Pathology report	
Other	Pulmonary function test		

2.3. Protocol alignment for "Part A" and "Part B"

Once protocols were created independently for the 4 disease cohorts, they were then reviewed iteratively by the overall team to identify commonalities between protocols. Common data types were grouped and a common protocol across all disease cohorts was created for "Part A". This part of the protocol was meant to include demographics, past medical history, confounders, acoustic tasks and patient-reported outcome measures (PROMs) that are valuable across all cohorts. Moreover, tasks used to screen for certain diseases were included. For example, the GAD-7 questionnaire was included in PART A as it is a common screening tool for anxiety. Therefore, a patient with a voice disorder doing the "voice disorder protocol" would still be screened for anxiety with the GAD-7 which could become a confounder in the analysis. A "Part B" was then developed with more specific questions or tasks that are particular to a disease cohort. For example, the "Part B" of the neurological disorder includes more in-depth neuro-psychological assessments.

2.4. Validation of the protocols for Standards, Ethics, Clinical expertise, Tools and PEDP

Once the draft "Part A- Across all cohort" and "Part B- Cohort Specific" protocols were developed, each section was reviewed by a group of eight experts during 3 different three-hour review sessions. Reviewers included clinical experts, bioethicists, standards experts, diversity experts, and software developers. Each question of the protocol was reviewed and amended based on the following determinations:

- Standards experts assessed if standard version of the question was available (ex: how to ask for gender, or socio-economic status based on Canadian and US Standards)
- Ethics experts assessed if there were any ethical issues with the question (ex: some questions about health history were asked before consent which was deemed unethical and had to be moved to after the consent section.
- DEI experts assessed if diversity and inclusion was well represented in each question (ex: some questions about disabilities were altered to focus on accessibility instead of disability status which was deemed to be discriminatory)
- Tools experts assessed how software development and technology could automate or facilitate data collection.
- Clinical experts assessed the clinical significance and validity of each task and rated them in order of clinical importance

3. Results

Protocols for data acquisition were harmonized across the 4 adult cohorts and a final protocol containing "PART A- across all cohorts" and "PART B- Cohort Specific" was created. Feasibility studies were conducted to assess feasibility metrics such as time of completion, understandability of tasks, task complexity, and need for assistance, and will be described in a separate manuscript. Overall, the mean protocol completion rate ranged between 45 and 75 minutes depending on the disease cohort, age of participant, and cognitive function.

The complete protocols include 51 demographic questions, 92 confounder questions, 22 acoustic tasks, and 12 PROs and validated questionnaires.

Full protocols will be available in the REDCap Instrument Shared Library and are also available for download on our GitHub repository https://github.com/eipm/bridge2ai-redcap.

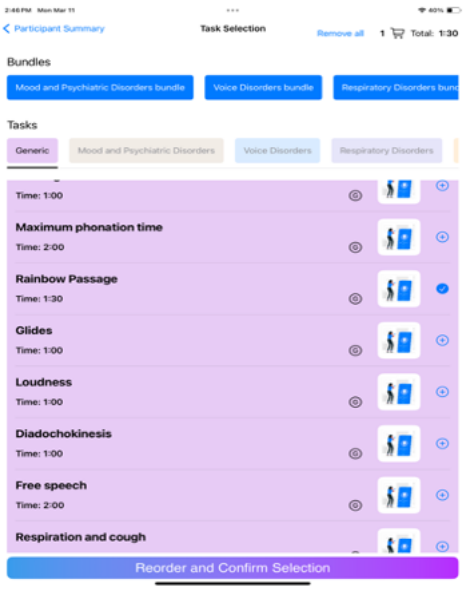
Table 2: Different Acoustic tasks included in the Bridge-2AI-Voice Protocol

Task	Section
Respiratory sounds	A, Bresp
Voluntary Cough	A, Bresp
Prolonged vowels	A
Maximum Phonation Time	A
Glides	A
Loudness	A
Diadochokinesis	A
Rainbow Passage	A
Caterpillar passage	Bvoice
Cape-V Sentences	Bvoice
Free Speech	A, Bvoice
Picture Description	A
Story Recall	A
Animal Fluency	Bmood
Open-ended questions	Bmood
Word-Color Stroop	Bneuro
Productive Vocabulary	Bneuro
Random item generation	Bneuro
Cinderella story	Bneuro

3.1 Development of protocols for the "Control Cohort"

After an extensive literature review on what represents a control cohort in voice and speech science, our group decided to avoid the term "normal voice and speech" when describing the control cohort but rather describe it as individuals that did not have any of the diseases studied in the "disease cohorts". This decision was made to avoid overfitting and optimize the generalizability and scalability of potential machine learning models trained on the datasets produced by these protocols. The control cohort protocol consisted of 1 of the protocols of the disease cohorts paired with a "control cohort" clinical validation section (see Annex 6).

Figure 1: Example of task selection tool



4. Discussion

Here, we have described the Bridge2AI-Voice protocols, a funded multi-institutional and multidisciplinary effort towards standardizing voice, speech, and respiratory sounds data collection protocols for broad applications within clinical sciences. Such an endeavor is paramount for voice biomarkers to be validated for disease monitoring, and establish a foundation for reliability and responsiveness studies using such biomarkers. To date, such efforts have largely been absent, with a few contemporary exceptions in specific disease categories [9], such as AD/ADRD, also associated with significant and commensurate funding.

One unique feature of the collective iterative process described in the methods is its multidisciplinarity, with the inclusion of acoustics scientists, diverse and expert clinicians, interoperability experts, mobile health developers, computer scientists, bioethicists, diversity, equity and inclusion specialists over several months. Such a colossal enterprise would not have been possible without the appropriate funding and underscores the importance of commensurate funding for moonshot programs to make the safe, ethical, and effective application of AI in healthcare a possibility. The iterative nature of the clinical protocol design, including the complex informed consent it relies on, has permeated the collaboration and is continuing to strengthen our approach as we learn new limitations from active patient recruitment and data collection. Another unique feature of our data collection protocol is the presence of a common trunk of data acquisition transcending specific disease categories, “PART A- across all cohorts". This common protocol was carefully crafted through multiple rounds of multidisciplinary discussions to include potential confounders such as demographic factors or voice-affecting conditions, such as occupational history, and to screen for possible diseases in our specialized cohorts, which may impact data labeling. In “Part B- Cohort Specific,” we focus on detailed voice, speech, and respiratory tasks relevant to the specific disease categories we are collecting data for. This bi-partite approach has the advantage of creating data type commonalities across cohorts, allowing for careful control cohort determination with explicit confounders identification. To date, this data acquisition strategy is unique in the rapidly evolving field of Audiomics.

Other unique features of our data collection protocols include its explicit open-access nature. Our multidisciplinary collective is focused on the good of humanity and aims to model good practice in science. Following the Open Science Framework [10], we are making the protocols transparent, open access, and widely available, and we welcome feedback for continuous improvement. We believe the desire to establish state-of-the-art protocol standards for voice AI biomarkers can only be accomplished through iterative revisions, as crystalized in our early feasibility work through early mobile app development and recruited subject feedback. Additionally, bioethics and DEI experts on the team critically review the consent and data access issues related to privacy, security, and potential discrimination and other harms. To date, the consent has undergone several rounds of revision. We are also exploiting several models of licensing and different level of access to optimize the safety of participating subjects while expanding research opportunities linked to data access.

From a technical standpoint, our team has made important and unique contributions that we attribute to the dynamic leadership, collaborative structure and significant multi-year funding mechanism supporting us. An output of particular impact is the incorporation of RedCap and electronic health record (EHR) interoperability standards based on FHIR from the outset of the data acquisition tool creation, again, which provides a pathway for establishing reusable standards for integrating voice information into EHR data. The sensitivity of this connected data has been primordial, and our early efforts on using federated learning may eventually combine with FHIR (e.g. (McMurry et al. 2024)) to provide privacy-preserving solutions to both data governance and computing. From an acoustics standpoint, our team has pushed forth a low-cost solution that is adaptable to the protocols for data acquisition for different disease categories, with minimal hardware adjustment. An innovation that is forthcoming is the incorporation of automated gain adjustment via our mobile app, to minimize data collector burden and ease potential future data acquisition in the wild with remote consent and no researcher involved.

5. Conclusions

For too long, voice/speech and respiratory sounds biomarker research has been limited by small-scale studies, usually based on single-institution data, with limited transparency regarding data collection protocol. From the voice tasks used to details on recruited populations to hardware specification, the lack of openness has stifled progress and limited the possibility of Audiomics with the rise of AI. The iterative, multidisciplinary, and collaborative approach of the Bridge2AI-Voice project in the formulation and continuous improvement of data acquisition protocols for Audiomics has led to several unique innovations. Herein we share our protocols widely with the scientific community, in the spirit of the Open Science Framework. Rather than stifling the pace of scientific progress, we strongly believe that transparency, continuous feedback, adoption of ethical principles, and technical meticulousness accelerate science and provide unique opportunities for innovation. Our efforts are in continuous expansion, as the desire to collaborate rather than compete with our collaborative continues to grow.

6. Acknowledgements

We would like to acknowledge the following Bridge2AI-Voice Consortium collaborators for their contribution: Kathy Jenkins, MD, Jessily Ramirez MBA, Daniel Low PhD, Jamie Toghranegar CCC-SLP, Gaetane Michaud, MD, Cindy Kostelnik, Ariel Farb, Desiree McCutcheon, Samantha Salvi-Cruz. All other collaborators that have contributed to this work have been named in the authorship line.

The Bridge2AI-Voice Consortium is funded by the NIH’s Common fund, Grant Number: OT2-OD032720- 01S1 and OT2-OD032720- 01.

7. References

[1] J. Robin, J. E. Harrison, L. D. Kaufman, F. Rudzicz, W. Simpson, and M. Yancheva, "Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations," *Digit Biomark*, vol. 4, no. 3, pp. 99–108, Sep. 2020.

[2] R. Fusaroli, A. Lambrechts, D. Bang, D. M. Bowler, and S. B. Gaigg, "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis," *Autism Res.*, vol. 10, no. 3, pp. 384–407, Mar. 2017.

[3] G. Fagherazzi, A. Fischer, M. Ismael, and V. Despotovic, "Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice," *Digit Biomark*, vol. 5, no. 1, pp. 78–88, Apr. 2021.

[4] M. Bowden, E. Beswick, J. Tam, D. Perry, A. Smith, J. Newton, S. Chandran, O. Watts, and S. Pal, "A systematic review and narrative analysis of digital speech biomarkers in Motor Neuron Disease," *NPJ Digit Med*, vol. 6, no. 1, p. 228, Dec. 2023.

[5] J. D. S. Sara, D. Orbelo, E. Maor, L. O. Lerman, and A. Lerman, "Guess What We Can Hear—Novel Voice Biomarkers for the Remote Detection of Disease," *Mayo Clin. Proc.*, vol. 98, no. 9, pp. 1353–1375, Sep. 2023.

[6] A. Idrisoglu, A. L. Dallora, P. Anderberg, and J. S. Berglund, "Applied Machine Learning Techniques to Diagnose Voice-Affecting Conditions and Disorders: Systematic Literature Review," *J. Med. Internet Res.*, vol. 25, p. e46105, Jul. 2023.

[7] D. M. Low, K. H. Bentley, and S. S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," *Laryngoscope Investig Otolaryngol*, vol. 5, no. 1, pp. 96–116, Feb. 2020.

[8] Y. Bensoussan, O. Elemento, and A. Rameau, "Voice as an AI Biomarker of Health-Introducing Audiomics," *JAMA Otolaryngol. Head Neck Surg.*, Feb. 2024.

[9] L. C. Kourtis, O. B. Regele, J. M. Wright, and G. B. Jones, "Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity," *NPJ Digit Med*, vol. 2, Feb. 2019.

[10] E. D. Foster and A. Deardorff, "Open Science Framework (OSF)," *J. Med. Libr. Assoc.*, vol. 105, no. 2, pp. 203–206, Apr. 2017.