MAYO CLINIC

MAYO CLINIC PROCEEDINGS

REVIEW

Guess What We Can Hear—Novel Voice Biomarkers for the Remote Detection of Disease

Jaskanwal Deep Singh Sara, MBChB; Diana Orbelo, PhD; Elad Maor, MD, PhD; Lilach O. Lerman, MD, PhD; and Amir Lerman, MD

Abstract

The advancement of digital biomarkers and the provision of remote health care greatly progressed during the coronavirus disease 2019 global pandemic. Combining voice/speech data with artificial intelligence and machine-based learning offers a novel solution to the growing demand for telemedicine. Voice biomarkers, obtained from the extraction of characteristic acoustic and linguistic features, are associated with a variety of diseases and even coronavirus disease 2019. In the current review, we (1) describe the basis on which digital voice biomarkers could facilitate "telemedicine," (2) discuss potential mechanisms that may explain the association between voice biomarkers and disease, (3) offer a novel classification system to conceptualize voice biomarkers depending on different methods for recording and analyzing voice/speech samples, (4) outline evidence revealing an association between voice biomarkers and a number of disease states, and (5) describe the process of developing a voice biomarker from recording, storing voice samples, and extracting acoustic and linguistic features relevant to training and testing deep and machine-based learning algorithms to detect disease. We further explore several important future considerations in this area of research, including the necessity for clinical trials and the importance of safeguarding data and individual privacy. To this end, we searched PubMed and Google Scholar to identify studies evaluating the relationship between voice/speech features and biomarkers and various diseases. Search terms included *digital biomarker*, *telemedicine*, *voice features*, *voice biomarker*, *speech features*, *speech biomarkers*, *acoustics*, *linguistics*, *cardiovascular disease*, *neurologic disease*, *psychiatric disease*, and *infectious disease*. The search was limited to studies published in English in peer-reviewed journals between 1980 and the present. To identify potential studies not captured by our database search strategy, we also searched studies listed in the bibliography of relevant publications and reviews.

 ■ Mayo Clin Proc. 2023;98(9):1353-1375

The use of telemedicine and digital biomarkers has increased over recent years. In this context, the coronavirus disease 2019 (COVID-19) global pandemic has provided a strong impetus for health care professionals to examine previously accepted practice patterns and to look for alternative ways of delivering health care. One consequence of this situation has been a dramatic shift from in-person care to non—face-to-face health care visits,[1] which has laid the foundation for remote clinics forming a more permanent aspect of health care delivery.[2-4] Studies have reported the benefits of remotely provided health care through event monitors, smart devices, and wearables on various disease processes including hypertension and heart failure.[5,6] Indeed, while traditional health care models require in-person evaluations with potentially lengthy visits and costly testing, telemedicine holds the promise of offering simple noninvasive methods of evaluation that can be undertaken remotely from health care professionals. Non—face-to-face clinical assessments also reduce the risk of

From the Department of Cardiovascular Medicine (J.D.S.S., A.L.) and Division of Otolaryngology (D.O.), Mayo Clinic College of Medicine and Science, and Division of Nephrology and Hypertension, Mayo Clinic (L.O.L.), Rochester, MN; Chaim Sheba Medical Center, Tel HaShomer, Israel (D.O.); and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel (E.M.).

ARTICLE HIGHLIGHTS

- Voice biomarkers offer promise in delivering health care remotely by providing adjunctive information that can aid health care professionals in making decisions about the detection, progression, and treatment of disease and could be used as surrogate end points in clinical trials.
- The association between voice biomarkers and systemic disease could relate to disease or injury affecting the lungs, vocal folds, or articulators (which produce phonation) or the brain and other neurologic structures (which influence phonation) or through systemic disease processes affecting the cardiovascular, endocrine, or immune systems, or could reflect our psychosocial states.
- Voice features and biomarkers are associated with a number of cardiovascular, neurologic, and psychiatric diseases and even coronavirus disease 2019.
- Classification of voice biomarkers involves outlining the source of vocalization, the method of voice assessment, the method of voice analysis, and the location of recordings.
- Developing voice biomarkers involves obtaining voice recordings, performing audio preprocessing, undertaking acoustic and linguistic feature extraction and selection, carrying out the training and testing of voice algorithms, and evaluating their integration into clinical practice.
- For the field of voice biomarkers to evolve, there is a need for large prospective clinical trials in demographically and linguistically diverse populations; special attention also should be given to defining how to safely integrate voice biomarkers into existing clinical and research work models in a way that prioritizes data protection and individual privacy.

transmitting infectious diseases and are more accessible, time-efficient, and cost-effective than face-to-face visits, which is advantageous to health care systems that are overloaded or resource-poor.

Voice/speech analysis is a promising noninvasive digital technology that can examine spoken language to provide information on health status and offers promise as a digital biomarker to screen for disease. In this article, *voice/speech data* refers to the sounds created by vocal fold oscillation, speech, and any other sounds created in a human vocal tract. Conversely, *language* is separate and is created by cognition and linguistic knowledge. Thus, language involves the production of speech sounds into meaningful units (ie, phonology), the combination of these meaningful units into words (ie, affixes and base words), and the generation of sentences and paragraphs to communicate thought.

While the vocal tract only produces voice and speech, cognitive and linguistic elements produce language. Generating voice/speech data encompasses a wide range of behaviors, from the production of a simple sound such as a sigh to natural language produced in conversation. Because voice/speech reflects the coordination of a variety of cognitive and motor processes, voice/speech could provide a sensitive snapshot of cognition, tissue integrity, and motor function relevant to many diseases. Indeed, studies have found that voice/speech contains acoustic and linguistic features that can be identified by machine learning models to gauge a speaker's behavioral health status.[7-11] Moreover, the collection of data for voice/speech analysis can be relatively simple and engaging for patients and is relatively inexpensive and convenient, requiring just a microphone, a quiet place, and a device to capture audio samples. Voice/speech samples can be collected remotely using a personal device and, with simple and clear instructions, does not require special training. Furthermore, voice/speech analysis permits repeated and frequent testing, reducing measurement error and providing richer and more detailed data across time. Thus, voice/speech—based biomarkers (henceforth referred to as *voice biomarkers*) could permit sensitive screening for disease, monitoring of progression and response to treatment and, with informed consent, could be a useful surrogate marker for clinical research.

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

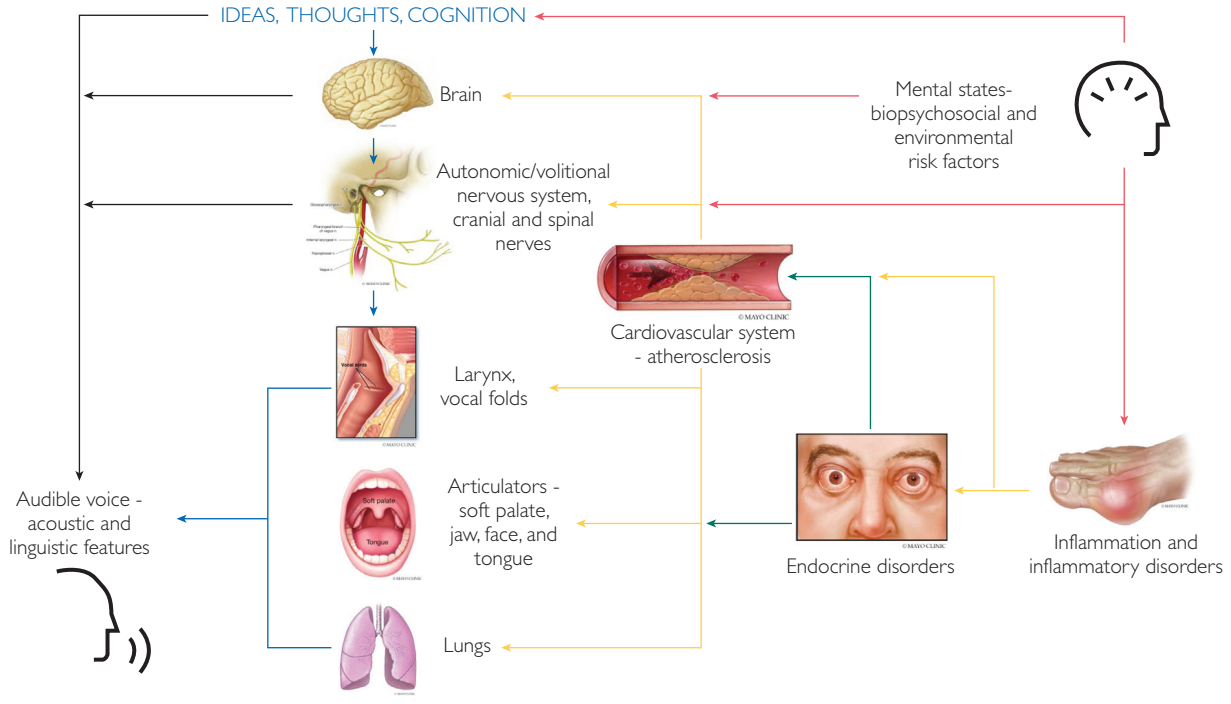

FIGURE 1. Classification of voice biomarkers with different active and passive methods for making voice recordings.

POSSIBLE MECHANISMS OF ASSOCIATION BETWEEN DISEASE AND VOICE FEATURES AND BIOMARKERS

Directly attributing specific voice features to disease processes would involve documenting the consistent replication of such features in the context of similar pathophysiologic findings, with concurrent evaluation of biologically plausible mechanisms. Treatment or progression of a given disease should correspond with predictable changes in voice features and the identification of reliable changes in the underlying pathophysiology that link the disease to the process of voice generation. The following sections outline our current understanding of possible mechanisms for the association between disease and voice features and biomarkers. For the field of voice/speech analysis–based disease detection to mature, large well-conducted studies elucidating associations and causal mechanisms will be required to help clinicians and researchers make more sense of these biomarkers.

Organs of Speech

Voice/speech generation is the expression of thoughts and ideas through the transformation of airflow into sound patterns. This process requires the cooperative efforts of the following structures: the lungs, which provide the air to create sound; the larynx, which contains the vocal folds whose vibrations create sound; and the vocal tract (the area from the top of the larynx to the lips that contains articulators including the tongue, jaw, and soft palate), which further shape sounds (Figure 1). Thus, there is the source (respiratory and laryngeal systems) and the filter (resonance and articulation systems). Voice/speech therefore refers to the production of sound emanating from air pressure and airflow generated by the respiratory and laryngeal systems, while the vocal tract is responsible for the process of resonation and the articulators can be positioned to influence resonance and produce speech sounds (ie, articulation). In other words, sounds come from the larynx, and

they are modified by the size and shape of the vocal tract. Abnormalities at any stage in this process may give rise to characteristic voice/speech features that may be detectable by machine and deep learning algorithms. These abnormalities could include aberrations to the normal structure and function of any of the component organs either from primary disease in organs of phonation or through secondary organ dysfunction such as central nervous system lesions or degeneration.

Neurocognitive and Neuromotor Function

Figure 1 outlines central contributors to voice/speech aberrations. Ideas form in our mind under the direction of the activity of a number of specific brain networks, such as the Broca and Wernicke areas in the left frontal and temporal cortices, respectively. Further, multiple cranial nerves participate in speech control, including cranial nerves V, VII, IX, X, XI, XII as well as the spinal nerves that control respiration. The vagus nerve in particular plays a critical role in phonation under the direction of the voluntary nervous system (recurrent laryngeal and motor branch of superior laryngeal nerves). Thus, disease or injury directly affecting these or related neurologic structures through processes such as ischemia, inflammation, trauma, neoplasm, and infection to name a few can provide a plausible basis for changes identified in voice/speech data.

Endocrine and Immune Systems

The link with disease in other systems, such as the endocrine and immune systems, is less well understood. Studies have found a link between systemic immune-mediated diseases, such as rheumatoid arthritis, Sjögren syndrome, and systemic lupus erythematosus, and voice disorders with significant correlations between biochemical parameters of inflammation such as sedimentation velocity and anti-DNA antibodies with detectable vocal changes.[12] Endocrine disorders such as diabetes mellitus[13] and hypothyroidism[14] have also been associated with specific voice/speech abnormalities, suggesting that the immunologic and hormonal milieu related to various systemic disorders may be expressed in voice/speech characteristics through secondary effects on organs of phonation. In one study, vocal cord and lung edema were associated with characteristic perturbations in vocal characteristics[15] that may be underpinned by transudative or exudative causes, the latter of which could be caused by inflammatory disorders.

Cardiovascular System

Cardiovascular disease (CVD) may similarly modify phonation through systemic processes. Atherosclerosis and vascular disease may lead to chronic hypoperfusion of the larynx and other organs of phonation or to the brain and other neurologic structures that influence voice production. This process may in turn lead to structural atrophy, involution, calcification, stiffening, and loss of structural integrity and muscle mass.[16] In the larynx, these changes could lead to incomplete closure of the vocal folds, reduced laryngeal mobility, reduced cartilage sliding, and decreased mucosal wave vibration, and with them, predictable changes in vocal characteristics such as an increase or decrease in fundamental frequency of phonation[16] among others. Moreover, most CVDs and their complications are caused by atherosclerotic disease, which is widely considered a systemic inflammatory process.[17] Thus, a specific cardiovascular inflammatory "soup" could exert unique secondary effects on the structure and function of organs of phonation, resulting in the expression of characteristic voice features. Cardiovascular disease may also play a role in modifying voice features through its intimate relationship with the autonomic nervous system. The vagus nerve innervates the heart through its superior, inferior, and thoracic cardiac branches and its activity is tightly coupled with heart rate control, which is in turn associated with coronary artery disease (CAD)[18] and cardiovascular events.[19] Indeed, both heart failure and

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

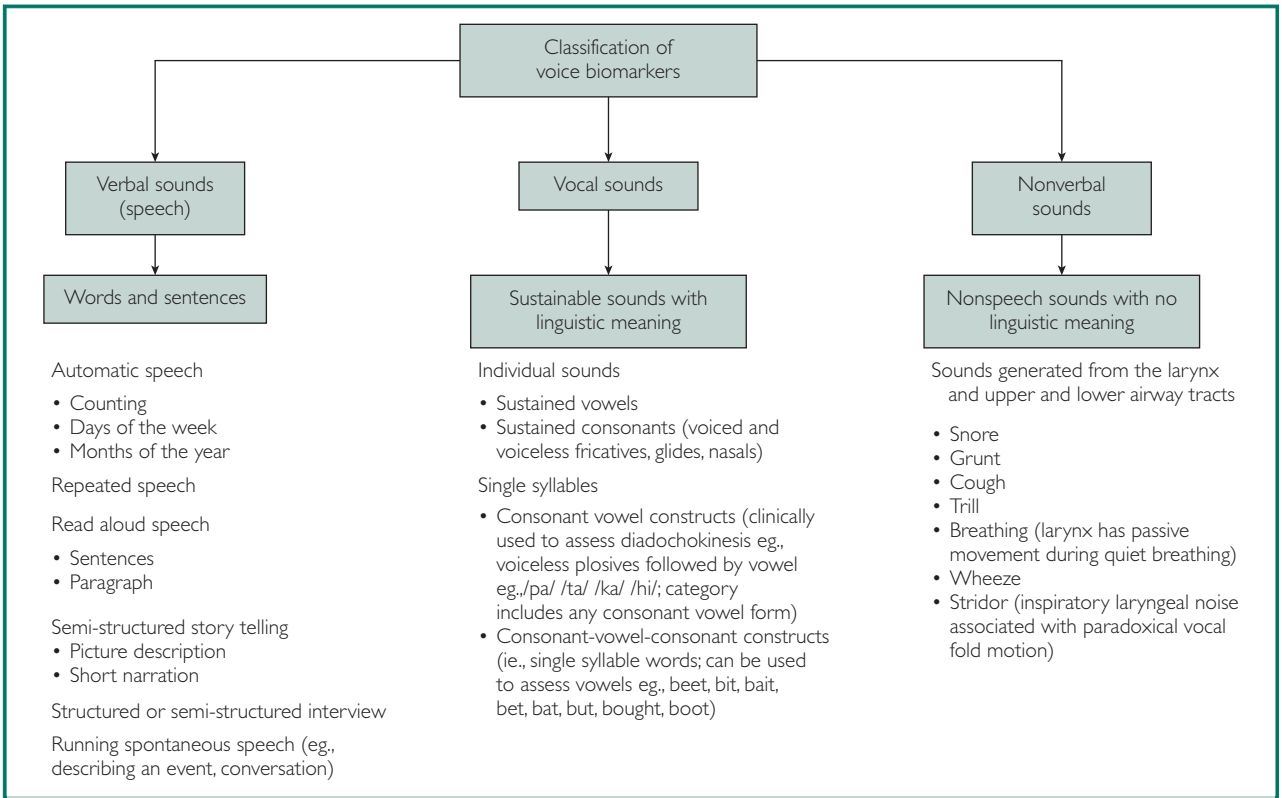

FIGURE 2. Outline of the potential underlying mechanisms that could explain the relationship between disease states and voice acoustic and linguistic features.

CAD are associated with dysregulation in autonomic nervous system activity manifesting with characteristic signs and symptoms of disease. Thus, it follows that similarly stereotypical changes in vocal/speech may also be evident given the key role of the key vagus nerve in phonation.

Mental States

Acoustic and linguistic features may represent an epiphenomenon occurring in parallel to disease rather than being directly caused by disease per se. For example, growing evidence supports the association between our mental states, and associated biopsychosocial factors, with disease. Mental stress in particular has been reported to be associated with diabetes mellitus[20] and CVDs including CAD, atrial fibrillation, and stroke.[21-24] Indeed, mental stress is associated with measurable changes in voice/speech characteristics,[25,26] and voice symptoms are common in individuals with high levels of cortisol.[27] Although these explanations would not account for the high variability of acoustic and linguistic features identified across different diseases, mental stress could still be playing a role in the voice/speech analysis of each of these disease states. Mental stress affects our internal dialogue through our cognitions and thoughts, which explains the unique linguistic patterns that can be seen in individuals with depression and other psychiatric diseases.[28,29] Mental stress is also increasingly being seen to directly affect the neurologic, endocrine, immune, and cardiovascular systems[30] through a variety of mechanisms including amygdala and sympathetic

nervous system overactivity, up-regulation of the hypothalamic-pituitary-adrenal axis, enhanced release of peripherally circulating inflammatory cytokines, and greater vascular inflammation, endothelial dysfunction, atherosclerotic disease progression, and cardiovascular events.[31] Thus, mental stress could be a contributing factor to the mechanistic pathways linking a variety of disease states with detectable voice/speech characteristics. More studies will be required to better clarify this and other potential causative mechanisms as we learn more about the potential role of voice biomarkers in clinical practice.

VOICE CLASSIFICATION

Currently, there is no practical framework for the conceptualization of voice biomarkers that distinguishes the source or characteristics of voice/speech data that can be analyzed and developed into potential biomarkers. These biomarkers may be derived from verbal sounds (speech), vocal sounds, and nonverbal sounds with active or passive tasks (Figure 2). On this basis, we propose a novel 4-position classification system for the development voice biomarkers. Position I refers to the type of vocalization used for analysis and extraction of a voice biomarker, with S denoting verbal (speech), V denoting vocal, N denoting nonverbal sounds (noise), and D denoting all. Position II refers to the assessment strategy used with A denoting voice recordings during active tasks and processes, P denoting passive voice recordings, and D denoting both. Position III refers to the analysis strategy with A denoting analysis of acoustics, L denoting analysis of linguistics, and D denoting both. Position IV refers to recordings performed in person, denoted by I, or remotely, denoted by R (Table 1).

TABLE 1. Proposed Novel Classification System for Voice Biomarkers

I	Source of vocalization	S — speech V — vocal sounds N — nonverbal sounds D — all types
II	Method of voice assessment	A — active recording P — passive recording D — both types
III	Method of voice analysis	A — acoustic L — linguistic D — both types
IV	Location of recording	I — in person R — remote

Table 2 further elaborates on this classification by providing examples of different methods of voice/speech analysis and different locations and modes of obtaining recordings. "Active" recordings require engaging participants in specifically outlined tasks, whereas "passive" recordings involve free and spontaneous speech and sounds. In our previous studies, individuals were asked to obtain 3 separate 30-second voice recordings—reading a prespecified text, describing a positive emotional experience, and describing a negative emotional experience.[32-34] This partially structured and semi-spontaneous approach provided the opportunity to assess a combination of acoustic (audible and nonaudible sound-related) and linguistic (semantic content—related) features; this offers a balance between "controlling" the vocal task and allowing participants to formulate their own responses. We have previously documented that 2 features derived from machine learning were independently associated with CAD. These features were (1) a statistical operator describing extreme values of the Mel Frequency Cepstral Coefficient (MFCC), which measures craters around specific frequency bands, and quantifies the width and depth of these craters, and (2) a skewness measure that portrays the asymmetry of intensity distribution in a specific frequency band.[32] Both were more strongly associated when participants were asked to describe a negative emotional experience.[32] Hence, unique voice/speech characteristics emerge in the acoustics and/or linguistic content produced by patients when articulating emotively significant ideas. Special attention should be given to the difficulty in obtaining spontaneous speech samples. Patients can often be reluctant to share positive

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

TABLE 2. Examples of Different Methods of Voice Analysis and Different Location and Modes of Voice Recordings

Variable		Examples	Potential clinical application
Method of voice analysis	Linguistic analysis	• Phonetic and phonological (eg, number of pauses, total pause time, hesitation ratio, speech rate) • Linguistic prosody (eg, melody of speech conveying attitude and emotion, high and low formants) • Lexicosemantic (eg, average rate of occurrence for each part of speech, number of repetitions, semantic errors, and closed-class word errors) • Morphosyntactic and syntactic (eg, number of words per clause, number of dependent and simple clauses, number of clauses per utterance, mean length of utterances) • Discourse-pragmatic (eg, cohesion, coherence)	Linguistic features extracted from spontaneous or semispontaneous speech for the detection of Alzheimer disease, mild cognitive impairment, or mental health disorders (eg, depression and posttraumatic stress disorder)
	Acoustic analysis	• Attitudinal and affective prosody (eg, frequency over time) • Fundamental frequency-based measures (eg, zero-crossing rate, harmonic-to-noise ratio, noise-to-harmonic ratio, jitter, and shimmer) • Spectral characteristics (eg, spectral flux, slope, centroid, entropy, roll-off, and flatness) • Phonation (eg, pitch period entropy) • Nonlinear dynamic features (eg, correlation dimension, fractal dimension, recurrence period density entropy, Lempel-Ziv complexity) • Segmental features (eg, MFCCs, perceptual linear prediction coefficients, linear frequency cepstral coefficients)	Acoustic features extracted from sustained vowel phonations or diadochokinetic tasks used in the detection of Parkinson disease or from individuals actively coughing and/or passively breathing used in the detection of COVID-19 pneumonia
Location of recording	In person	• Studio- or clinic-based	Controlled environment where unwanted noise is reduced, requires calibrated equipment, prevents remote testing, requires professional staff (eg, sound engineer)
	Remote	• Telephone-based	Allows remote testing, involves variable equipment, restricted recording bandwidth, and does not permit environment control
		• Online computer–based	

Continued on next page

TABLE 2. Continued

Variable	Examples	Potential clinical application
		Accessible and widely available, permits large-scale data collection
	• Smartphone-based	Widely available, low cost, potential for greater bandwidth than telephone recordings, permits remote and frequent monitoring "on-the-go," does not permit environment control

COVID-19, coronavirus disease 2019; MFCC, Mel Frequency Cepstral Coefficient.

or negative emotional experiences and should be encouraged to expand on what they are saying. Spontaneous speech samples are valuable when performing voice/speech analyses but are more challenging to obtain than reading passages or sentence repetition. The acoustic signal may also vary more in spontaneous conversational speech compared with controlled vocalization tasks (eg, sustained vowel sounds). With more voice disruptions and quality fluctuations,[35] spontaneous speech may be better at eliciting changes in the organization and complexity of language. Conversely, reading scripted text may capture important changes in acoustic characteristics of speech because speakers tend to be more fluent in read speech. Reading tasks provide more control over the speech task but require instruction and active engagement. Consideration must also be given to individual reading skill. Moreover, sustained vowel phonations can provide information useful for evaluating dysphonia and assess the integrity of the laryngeal and respiratory systems while eliminating the influence of confounders such as speaking rate, intonation, dialect, accent, and even language.[35] Conversely, diadochokinetic tasks (fast repetition of syllables, eg, /pa/ta/ka/) require rapid movements of the jaw, lips, and tongue, challenging individuals to maintain the rate and clarity of speech. These tasks provide the greatest level of control over a task, do not require complex instructions, but still require active patient participation and provide information for only a narrow range of conditions. Thus, the selection of approach for voice/speech recordings and analysis may depend on several factors including resources available and the objective of testing because different types of voice/speech samples assess the mechanism of voice and speech generation in different ways. For example, if the purpose is to screen for a psychiatric condition, in which the content and cadence of speech may be characteristic for specific disorders like depression, or when screening for cognitive impairment in an individual at risk for Alzheimer disease, in which language content, flow, and semantic errors may provide important diagnostic clues, then linguistic analyses maybe more relevant. Conversely, if the goals are to monitor treatment response in individuals with Parkinson disease in whom variations in the quality of voice and articulation are expected, then analyzing acoustic features may be more helpful.

VOICE/SPEECH ANALYSIS ASSOCIATED WITH DISEASES

Studies have found that features identified from voice/speech data are associated with several disease states (Table 3). An important distinction to note is that *voice features*

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

refers to individual acoustic (audible and nonaudible sound—related) and linguistic (sematic content—related) characteristics that can be identified from voice/speech data using machine learning. A voice *biomarker* is typically identifiable when more than one of these unique features is integrated into a single combined measure. Although single-feature voice biomarkers are possible, in this article we refer to *voice biomarkers* as reflective of multiple voice/speech features. We searched the PubMed and Google Scholar databases to identify studies evaluating the relationship between voice/speech features and biomarkers and various diseases. Search terms included *digital biomarker*, *telemedicine*, *voice features*, *voice biomarker*, *speech features*, *speech biomarkers*, *acoustics*, *linguistics*, *cardiovascular disease*, *neurologic disease*, *psychiatric disease*, and *infectious disease*. The search was limited to studies published in English in peer-reviewed journals between 1980 and the present. To identify potential studies not captured by our database search strategy, we also searched studies listed in the bibliography of relevant publications and reviews.

TABLE 3. Examples of Diseases Associated With Features and Biomarkers Derived From Voice Signal Analysis

Cardiovascular diseases
• Coronary artery disease at baseline
• Coronary artery disease events at follow-up
• Pulmonary hypertension
• Heart failure hospitalization and mortality
Neurologic diseases
• Parkinson disease
• Alzheimer disease
• Mild cognitive impairment
Infectious diseases
• Coronavirus disease 2019
Psychiatric conditions
• Autistic spectrum disorder
• Posttraumatic stress disorder
• Major depression
• Bipolar disorder

Cardiovascular Disease

To our knowledge, the first study assessing the relationship between acoustic characteristics in voice with cardiovascular health was published in 1990.[16] In that study, 18 nonsmoking vocally untrained males divided into groups of healthy young and healthy elderly men and elderly men who had chronic atherosclerosis without other systemic complaints were asked to enunciate the sound *a* at a manageable pitch sustained at 70 to 78 dB. Thus, according to our proposed classification system for voice biomarker analysis (Table 1), this study relied on a vocal active acoustic in-person (VAAI) biomarker. Significant differences between groups included mean fundamental frequency, the standard deviation of amplitude, percent jitter (cycle-to-cycle variations in vocal fundamental frequency), and shimmer (cycle-to-cycle variations in vocal amplitude). These differences were most apparent when comparing healthy young males to elderly males with atherosclerotic disease, and although only percent jitter significantly differentiated the 2 elderly groups, those with atherosclerotic disease generally exhibited phonation with greater short- and long-term variability. These findings suggested that the elderly speakers may be more prone to detectable disruptions in vocalization when diseases are present that have the potential to disrupt vocal fold vibration.[16] This study provided one of the earliest indications that voice/speech analysis could indicate the presence of systemic pathology.

Our group has extended this work by further evaluating the association between voice biomarkers and CVD[32-34,36] (Table 4) using a speech active acoustic in-person (SAAI) biomarker. Patients presenting to our institution for clinically indicated coronary angiography performed 3 separate 30-second speech recordings on an iPad or smartphone prior to their procedure. In the first of our studies, we used the MFCC to extract information from the speech recordings,[37] which is a sound processing tool used to classify healthy and

abnormal voices.[38-40] Speech recordings were then further analyzed using the Fourier transform mathematical function to convert data into patterns of frequency over time.[41] We identified and extracted 81 prespecified voice features from each participant's recordings. In a multivariate binary logistic regression analysis, we identified 2 features that were associated with CAD.[32] In our subsequent studies, we used a single integrated voice biomarker that was a composite of a total of 223 acoustic features, which included the Mel Cepstrum representation, jitter, shimmer, loudness, and pitch. This biomarker was developed and trained in a large cohort of patients with chronic medical conditions, including congestive heart failure, who were registered to a health care call center.[36] Low-level acoustic features were presented in a temporal resolution of 100 points per second producing a matrix of 2000 columns, with the number of extracted temporal features as rows. High-level features were then extracted from these matrices using a moments analysis, and a machine learning model was employed to generate the vocal biomarker, which was optimized to the training cohort using cross-validation techniques. In a cohort of more than 2000 patients in Israel, the voice biomarker was associated with all-cause mortality and heart failure hospitalization at follow-up. When we used the same biomarker in patients who presented to the catheterization laboratory for clinically indicated right-sided heart catheterization procedures, we found that the biomarker was significantly associated with mean pulmonary arterial pressure and pulmonary vascular resistance. Most recently, we reanalyzed the speech samples of the cohort of patients who underwent coronary angiography in our first study using the integrated voice biomarker and prospectively monitored these individuals for incident CAD events. In a multivariable analysis, patients with a high voice biomarker value had a significantly higher risk of incident coronary disease events, including being admitted to the hospital with chest pain and an acute coronary syndrome (Table 4). The studies underscore the association between voice/speech analyses, based on acoustic features, and a range of different CVDs.

Neurologic Disease

Parkinson disease currently lacks a definitive diagnostic test, and its detection often requires examination by a movement disorder specialist. Parkinson disease is associated with well-characterized effects on speech, such as impaired phonation (hypophonia) and articulation (hypokinetic dysarthria), reduced energy in the higher parts of the harmonic spectrum and variations in pitch, and a high prevalence of voice disorders.[42] Thus, a digital voice biomarker could provide an objective tool to identify and even "quantify" symptoms and assess response to treatment. Studies have found that voice dysfunction may sometimes be an early sign of motor impairment in Parkinson disease with perturbations in the fine motor control of vocalization and may occur before less subtle changes in the limbs.[43] Thus, it follows that biomarkers evaluating verbal and vocal sounds obtained during active or passive tasks using an analysis of acoustic features could provide useful information when evaluating Parkinson disease. In one study, the investigators analyzed a database of speech recordings of patients with Parkinson disease and individuals without Parkinson disease using machine learning models to predict disease with a speech passive acoustic remote (SPAR) biomarker. Their machine learning model could discriminate patients with Parkinson disease from healthy controls even among those with few to no symptoms. These findings reveal that voice/speech analysis could facilitate the early and rapid detection of Parkinson disease,[44] potentially in the initial stages

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

TABLE 4. Summary of Findings of Studies Evaluating the Association Between Voice Biomarkers and Cardiovascular Disease Outcomes

Reference, year	Sample size and description	Age (y)	Male, No. (%)	Follow-up	Method of voice sample collection	Voice biomarker used	Cardiovascular disease outcome	Principal findings
Maor et al,[32] 2018	101 Patients undergoing planned coronary angiography and 37 controls in Rochester, MN Control group: 15 patients undergoing noncardiac procedures (eg, hernia surgery) and 22 undergoing no procedures	Median, 61; IQR, 51-67	54 (54)	NA (cross-sectional study)	In person, using a smartphone or iPad, 3 separate 30-s voice recordings (R1—reading prespecified text; R2—describing a positive emotional experience; R3—describing a negative emotional experience)	MFCCs used to extract information from the recordings with further analysis using the Fourier transform mathematical function; this allowed the extraction of 81 different prespecified voice features from each recording	Presence of CAD on invasive coronary angiography	Univariate binarylogistic regression analysis identified 5 voice features that were associated with CAD (P<.05 for all) Multivariate binary logistic regression with adjustment for ASCVD risk score identified 2 voice features associated with CAD (OR, 95% CI: 0.37, 0.18-0.79; and 4.01, 1.25-12.84;P=.009 and P=.02, respectively) Both features were more strongly associated with CAD when patients described an emotionally significant experience
Maor et al,[36] 2020	2267 Patients with CHF	Median, 77; IQR, 68-83	1434 (63)	Median, 20 mo; IQR, 9-34 mo	Analysis of anonymized audio phone recordings of patients	A total of 223 acoustic features were extracted from 20 s of speech for each patient The voice biomarker was developed on a training cohort of patients who did not have CHF (N=8316)	Primary outcome: all-cause mortality Secondary outcome: hospitalization during follow-up	824 Patients(36%) died; 1403 (62%) were admitted for at least 24 h at follow-up Kaplan-Meier survival analysis showed higher cumulative probability of death with increasing voice biomarker quartiles (23%, 29%, 38%,

Continued on next page

TABLE 4. Continued

Reference, year	Sample size and description	Age (y)	Male, No. (%)	Follow-up	Method of voice sample collection	Voice biomarker used	Cardiovascular disease outcome	Principal findings
								and 54%;*P*<.001); there were 313 (54%), 335 (58%), 357 (63%), and 398 (70%) hospitalizations in the 4 biomarker quartile groups (P<.001) In multivariable analysis, compared with the lowest quartile, patients in the highest quartile (Q4) were 96% more likely to die (*P*<.001) and 50% more likely to require hospitalization (*P*<.01)
Sara et al,[33] 2020	83 Patients undergoing clinically indicated invasive cardiac catheterization studies	Mean ± SD, 61.6±15.1	37 (45)	NA (cross-sectional study)	In person, using a smartphone or iPad, 3 separate 30-s recordings (as in Maor et al[32])	The vocal biomarker used was the same biomarker developed in the cohort of patients with CHF (Maor et al[36]) Intraclass correlation coefficient between the separate voice recording biomarker values was 0.829 (95% CI, 0.740-0.889)	Diagnosis of moderate or greater PH (defined as a mean PAP ≥35 mm Hg)	Patients with a high mean ± SD PAP had significantly higher values of the mean voice biomarker compared with a lower mean PAP (0.74±0.85 vs 0.40±0.88; *P*=.046) Multivariate logistic regression showed that an increase in the mean voice biomarker by 1 U was associated with a high PAP (OR.

Continued on next page

TABLE 4. Continued

Reference, year	Sample size and description	Age (y)	Male, No. (%)	Follow-up	Method of voice sample collection	Voice biomarker used	Cardiovascular disease outcome	Principal findings
								2.31; 95% CI, 1.05-5.07; *P*=.038)
Sara et al,[34] 2022	108 Patients undergoing planned coronary angiography in Rochester, MN	Mean ± SD, 59.5±11.4	59 (55)	Median, 24 mo; IQR, 1-60	In person, using a smartphone or iPad, 3 separate 30-s recordings (as in Maor et al[32])	The vocal biomarker used was the same biomarker developed in the cohort of patients with CHF (Maor et al[36])	Primary outcome: composite of presenting to the emergency department with chest pain and/or being admitted to hospital with chest pain and/or having an acute coronary syndrome Secondary outcome: composite of a positive stress test result at follow-up and/or the presence of CAD at follow-up coronary angiography	In multivariable Cox proportional hazards models adjusting for CAD grade on baseline angiography, a high baseline mean voice biomarker was associated with both the primary composite outcome (HR, 2.61; 95% CI, 1.42-4.80; *P*=.002) and the secondary composite outcomes (HR, 3.13; 95% CI, 1.13-8.68; *P*=.03)

ASCVD, atherosclerotic cardiovascular disease; CAD, coronary artery disease; CHF, chronic heart failure; HR, hazard ratio; IQR, interquartile range; MFCC, Mel Frequency Cepstral Coefficient; NA, not applicable; OR, odds ratio; PAP, pulmonary artery pressure; PH, pulmonary hypertension; Q, quartile; R, recording.

of disease when changes in voice are otherwise unrecognized. In another study, the investigators obtained voice samples from patients with Parkinson disease who were asked to prolong "ahhh." They analyzed 132 acoustic features to generate a vocal active acoustic remote (VAAR) biomarker. When using only 10 acoustic features, the authors documented almost 99% classification accuracy.[45] Additional studies have considered voice/speech analysis as a method to detect the progression of disease and its response to therapy. In one study, the investigators extracted features of Parkinson disease progression in 6000 speech recordings from patients with Parkinson disease to generate a speech active acoustic remote (SAAR) biomarker. The authors documented clinically useful accuracy in estimates of measures of disease progression when comparing clinician- and voice algorithm–derived estimates.[46] In another study using a VAAR biomarker, the authors found a significant correlation between the degree of improvement in voice quality attributed to therapy with levodopa and the extent of changes in voice quality assessed using their biomarker, suggesting its potential use in helping to target therapy more effectively.[47]

Changes in language ability may be apparent many years before patients or family members recognize any symptoms of cognitive impairment.[48,49] Indeed, Alzheimer disease, and cognitive impairment more generally, is known to affect spoken language in several ways, including alterations in verbal fluency reflected in hesitant speech and slow speech rate, word finding difficulties that result in the use of filler sounds such as "uh" and "um," circumlocutions, neologisms, grammatical simplification, semantic errors, and repetitions.[50] Thus, voice/speech analysis that incorporated an evaluation of speech or voice sounds with a focus on principally linguistic, with some acoustic, features would seem most relevant in this population. In one study, speech samples were obtained from individuals in whom Alzheimer disease was identified during life (and confirmed at postmortem examination), and each participant exhibited objective signs of mild cognitive impairment 6 to 18 months prior to development of overt Alzheimer disease. Speech samples were analyzed to develop a speech active linguistic in-person (SALI) biomarker. Subtle changes in speech patterns were noticeable even during the prodromal stages of Alzheimer disease, and linear trends were evident in successive stages of disease progression.[51] In an additional study that also developed and used a SALI biomarker, the authors documented significant differences in linguistic features, such as "hesitation ratio, speech tempo, length and number of silent and filled pauses, length of utterance," between healthy controls and adults with mild cognitive impairment while performing memory tasks.[52] Interestingly, acoustic features have also been used to screen for Alzheimer disease in its preclinical states[53] and to distinguish healthy individuals from those with mild cognitive impairment and Alzheimer disease[54] using an SAAI biomarker. These findings support the notion that voice/speech analysis could provide an objective means of discriminating pathologic cognitive changes from normal age-related changes. Another study used both linguistic features, measured from transcripts, and acoustic features, measured from audio files, to develop a speech active dual (linguistic and acoustic) in-person (SADI) biomarker. The biomarker achieved greater than 80% accuracy in discriminating individuals with and without Alzheimer disease.[55]

Voice/speech analysis has also provided useful information in children with neurodevelopmental disorders. In autistic spectrum disorders, deficits in social interaction are evident, and in those children who are able to speak, abnormalities in speech patterns and voice quality are often exhibited. However, the diagnostic utility of these findings is unknown, and they remain challenging to assess objectively. In a study evaluating speech samples of children with and without

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

autistic spectrum disorders using an SAAI biomarker, the investigators found significant differences between groups. Pitch variability yielded 90% specificity and 80% sensitivity in classifying autism in this sample.[56]

Infectious Diseases

COVID-19 is a systemic illness with primarily upper and lower respiratory tract symptoms. A suitable voice biomarker might evaluate acoustic features of nonspeech airway sounds such as breathing and coughing but potentially also speech production. Ideally, these samples could be obtained remotely to limit risk of disease transmission. One study used recordings of a forced cough over a cell phone to develop a nonverbal active acoustic remote (NAAR) biomarker, which was trained on 4256 individuals and tested on 1064. Based on formal microbiological testing for COVID-19, the biomarker achieved a sensitivity of 98.5% and a specificity of 94.2%, with an area under the curve (AUC) of 97%.[57] In another study, the investigators developed a VAAR biomarker in a training cohort of 434 voice samples in 272 participants, 160 (59%) of whom had COVID-19 on formal testing.[58] Transfer learning and adaptation methods were employed for feature extraction, and each 10-second recording of speech was converted to a Mel spectrogram and was passed through a pretrained convolutional neural network architecture. Two classification models, random forest and support vector machine, were then evaluated at different regularization levels and the best model was selected and tested in 80 individuals, 40 of whom tested positive for COVID-19. The voice biomarker differed significantly among infected compared with noninfected individuals, with an AUC of 72%. When the voice biomarker was added to self-reported symptoms, the AUC significantly improved to 85%.[58]

Psychiatric Conditions

Psychiatric evaluation often includes judgments of speech loudness, rate, and prosody (fluctuations in rate, pitch, and loudness that signal emphasis and emotion) as well as the organization, content, and meaning of language, which, taken together, convey information about emotional state, mood, and clarity of thought, all of which can contribute to psychiatric diagnosis. Posttraumatic stress disorder (PTSD) is a psychiatric condition whose diagnosis typically relies on interview and measures of self-report, which can lead to underdiagnosis and overdiagnosis. Given the nature of PTSD, both acoustic and linguistic features derived from speech are likely to be informative in evaluating this condition. One study analyzed speech samples obtained during clinical interviews from warzone-exposed veterans and included 52 cases of individuals with PTSD and 77 controls using an SADI biomarker. From 40,526 speech features that were input to a random forest algorithm, 18 speech features were ultimately used and collectively provided an AUC for a diagnosis of PTSD of 95.4%.[59] Another area in which clinical interviews and self-report measures are used and often lack accuracy is suicidal ideation, which was evaluated in US veterans again using an SADI biomarker. Fifteen acoustic and linguistic features derived from speech correctly classified suicidal ideation in veterans, with an AUC of 80%.[60]

Voice analysis has also been used to facilitate the detection of depression.[61] In one online study, investigators extracted features from anonymous speech samples of depressed individuals for depression severity

TABLE 5. Important Characteristics for Voice Features and Biomarkers

- Can be done remotely (over the phone or other digital device)
- Noninvasive and nonintrusive
- Acquired in a short period of time
- Independent of language and geographic location
- Independent of age
- Specific for disease groups
- Can be combined with other digital technologies
- Allow patient identification in a secure and encrypted way
- Can be integrated with established clinical protocols and practice patterns
- Can be integrated into the electronic medical record

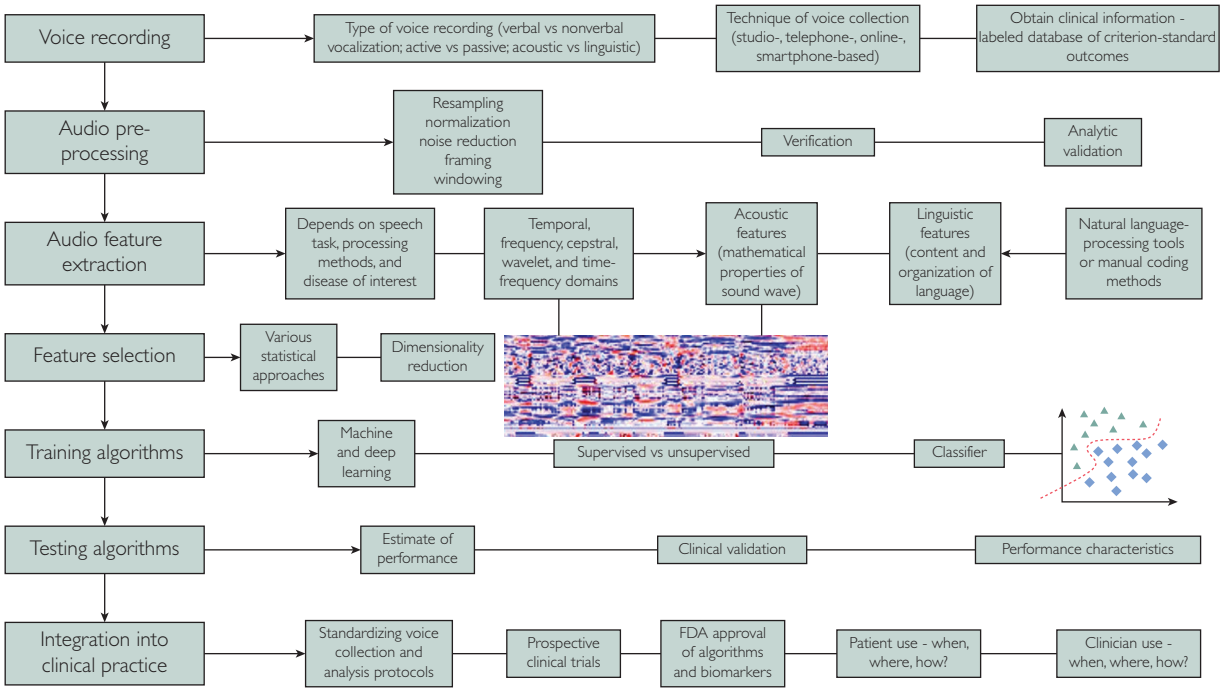

FIGURE 3. Outline of the process required for developing a voice biomarker for use in clinical practice. FDA, US Food and Drug Administration.

and suicidality using a speech active dual (acoustic and linguistic) remote (SADR) biomarker. Voice features predicted the Patient Health Questionanaire-9 scores with an AUC of 82.1%.[62] Similar results were obtained using a biomarker developed in person,[63] using linguistic assessment only during nonstructured interviews with research psychologists,[64] and using acoustic features only in individuals asked to read a standardized passage in a nearly closed room environment.[65] Interestingly, different acoustic features have been found to be helpful in both men and women, with spectral and energy-related features being helpful in discriminating depression in males (sensitivity of 0.95, specificity of 0.88) and prosody-related features being more helpful in females (sensitivity of 0.73, specificity of 0.86), suggesting the need for stratifying biomarkers across different demographic groups.[66] The SAAR biomarkers have also been used to monitor therapy in patients with depression, including in a double-blind trial in 105 adults with major depression who were randomized to sertraline (50 mg/d) vs placebo. Clinical changes in depressive symptoms among responders to pharmacotherapy were associated with significant differences in speech patterns that were not observed or were directionally opposite to changes seen in nonresponders.[67] Similar findings were observed in an observational study in which acoustic measures correlated significantly with baseline depression severity and were significantly different between treatment responders and nonresponders.[68] Lastly, acoustic features extracted from voice recordings made in naturalistic settings obtained using smartphones have also been used to discriminate bipolar disorder using an SADR biomarker.[69] Acoustic features

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

have also been used to distinguish depression from mixed depression and hypomania from mixed hypomania in individuals with bipolar disorder during an acute episode using an in-person verbal fluency task.[70] Thus, voice analysis technologies may increase the range of available methods that can be used to explore the neurobiological substrates of psychiatric conditions given their inherent dynamic interplay of systems that direct the motor, cognitive, and emotional basis of speech production.

DEVELOPING THE BIOMARKER

Table 5 lists important characteristics of voice biomarkers. There is currently no standard protocol for obtaining and analyzing voice/speech samples for the development of biomarkers. Figure 3 outlines potential steps for developing and using voice biomarkers in research and clinical practice.

Quality Control

Voice/speech data need to be recorded in conjunction with the collection of relevant clinical information to provide a reference data set of labeled criterion standards of outcomes to facilitate the training and testing of voice biomarkers. Audio recordings then must be preprocessed. This phase involves steps to improve the performance of feature detection: resampling allows improvement of poor-quality recordings and increases the number of high-quality recordings to reduce measurement error; normalization reduces extraneous information collected without altering differences in the ranges of acoustic values; noise detection and reduction generates a "clean voice estimation" through passing the "noisy voice" through a linear filter, and framing involves dividing the voice/speech sample into a number of samples, which are then multiplied by a "window function" to limit signal leakage effects.[71]

The Digital Medicine Society has provided a framework for consistent evaluation of voice/speech–based biomarkers,[72] which comprised 3 principal components: verification, analytic validation, and clinical validation.

Verification

Verification involves evaluating the quality of recordings and comparing the effects of recording devices, hardware, and sensors as well as recording conditions, including the effects of ambient noise, on the integrity of the recordings.[72] Layout of the user interface must also be considered, including the provision of instructions, whether there are empty periods in the recordings to determine ambient noise, and whether feedback is provided using voice prompts or beeps to improve adherence to the task. Ideally, the suitability of conditions can be communicated to the user via intrinsic calibration tasks to ensure that the quality of recordings are sufficient for accurate analysis.

Analytic Validation

Analytic validation includes assessing the validity and reliability of voice/speech sample processing and includes reproducibility and demographic dependence (eg, age, education, diagnosis, and accent). Analytic validation also involves comparing voice/speech features against reference standards.[72] The processing steps involved can often be complex because the features that can be extracted from voice/speech are highly variable and fluctuate according to the task.

For feature extraction, features are defined as "the most dominating and discriminating characteristics" of a speech/voice sample[71] and ultimately contribute to the training of machine learning algorithms for the detection of disease. The type of features used to generate a particular voice biomarker depends on the disease of interest and the recording and processing methods available. These include acoustic features that reflect mathematical properties of the sound wave (Table 2). Many acoustic features are computed via the mathematical transformations of the sound wave and can thus be validated using mathematical models

TABLE 6. Key Areas of Further Investigation and for Future Consideration in Developing Vocal Biomarkers in Health Care Delivery

- Greater standardization in voice data collection and analysis protocols
- Generating large-scale voice sample repositories labeled with clinical outcomes
- Systematic and rigorous evaluation of voice biomarkers in demographically and linguistically diverse populations from different geographic regions
- Large prospective clinical trials evaluating the use of voice biomarkers as surrogate measures of disease
- Voice biomarkers integrated into clinical practice—at-home monitoring using personal device, with feedback and other features
- Individuals using voice biomarkers, among other digital biomarkers, at home in smart devices such as smart mirrors
- Agency approval (eg, US Food and Drug Administration)
- Medical device and biomarker companies' long-term business models
- Insurance companies paying for the use of voice technology in clinical practice
- Voice biomarker as a proxy secondary end point in research studies
- Caution about false-positives and false-negatives and complacency of relying excessively/exclusively on voice biomarkers
- Lack of established guidelines outlining best practice for remote care
- Potential for unintended consequences from the "digital divide"—both positive and negative
- Understanding the mechanism linking changes in voice signal to disease states
- Using artificial intelligence–derived output in clinical practice without understanding "what's under the car's bonnet"
- Data protection and privacy—need for regulation and sophisticated encryption methods

of voice/speech production. The MFCCs are the most frequently used acoustic feature extracted and used in speech analysis, followed by perceptual linear prediction coefficients and linear frequency cepstral coefficients.[32] Acoustic features can be identified by applying the temporal, frequency, cepstral, wavelet, and time-frequency domains.[73] Conversely, linguistic features, reflecting the content and organization of language (Table 2), can be obtained using natural language-processing tools or manual coding methods. These techniques require prolonged speech recordings to ensure that features may be extracted at all linguistic levels. Features must then be "selected," for which various statistical techniques can be employed including minimum redundancy maximum relevance[74] or Gram-Schmidt orthogonalization.[75] These techniques allow a population of features to be selected without alteration by removing highly correlated features and those with missing values or low variance. Selected features can then be used individually or in various combinations as a composite biomarker to train machine and deep learning algorithms to classify clinical outcomes of interest.[71]

Clinical Validation

Clinical validation verifies the correspondence of a voice biomarker to a suitable clinical reference standard or several surrogate measures. In clinical validation studies, investigators should report clinically useful performance metrics such as accuracy, specificity, sensitivity, precision, and AUC to guide clinical application. When voice biomarkers are used, if they comprise multiple aggregated voice features, it is useful to also determine which individual features contribute most to classification models to help with clinical interpretation and to better understand the changes that accompany a disease. Lastly, if such voice biomarkers are found to be valid, reliable, and of potential clinical utility, additional steps will be required to ensure their safe and effective use in clinical practice (see subsequent section).

FUTURE DIRECTIONS

Table 6 summarizes key areas in voice/speech analysis that require further investigation. First, there needs to be greater standardization in voice/speech data collection and analysis protocols. Voice/speech recordings will almost exclusively be done outside a laboratory setting, in day-to-day situations and environments using an range of devices. Such standardization would permit cross-comparisons and transferability of files and data sharing. This step would pave the way for large-scale voice/speech sample repositories labeled with clinical outcomes, essential for ongoing research. Moreover, further clinical trials with large, demographically and linguistically diverse populations from different geographic regions assessing the use of voice biomarkers in evaluating a

Mayo Clin Proc. ■ September 2023;98(9):1353-1375 ■ https://doi.org/10.1016/j.mayocp.2023.03.007
www.mayoclinicproceedings.org

variety of disease states will be needed to ensure that they are providing universally reliable and valid measurements. This process of estimating and then improving upon the accuracy of voice biomarkers should be viewed as an iterative process rather than a conclusive outcome, particularly when considering how machine learning performs better with more input. Consideration should be given to the broader integration of voice/speech analysis technology into clinical practice and society. Voice biomarkers should be viewed as adjunctive tools that can aid physicians in making decisions while offering advantages of limiting unnecessary in-person evaluations that may be lengthy and expensive. In the modern era, telecommunications are by and large provided digitally, which coupled with the high proportion of individuals owning a smartphone or tablet provides the perfect alliance for the remote recording, transmission, and analysis of digital voice/speech recordings. One potential scenario is the adjustment of medications based on serial measurements of voice biomarkers across time as in patients with heart failure when titrating diuretics or other guideline-directed medical therapy. Another scenario could be the weekly or even daily *physical* done from home using voice biomarkers that could provide feedback to patients and health care professionals on clinical status. Voice biomarker algorithms embedded within medical devices would raise further questions about agency approval, such as from the US Food and Drug Administration; business models for device and biomarker developers; and whether insurance companies would provide coverage. In clinical research, if voice biomarkers prove to be valid and reliable markers of disease progression, they could potentially be used as cost- and time-efficient proxy end points in clinical trials. This feature would help make trials less expensive and allow more participants to be recruited and more end points to be studied and, specifically in pharmaceutical trials, could help identify drugs with (and without) promise sooner.

These exciting prospects should be viewed with caution, however. Relying on such home monitoring systems to prompt treatment decisions and whether patients should see their health care professionals makes ensuring the reliability and validity of voice biomarkers even more important. The potential adverse consequences of false-positives and false-negatives must be safeguarded against, further underscoring the need for large prospective studies evaluating the use of voice biomarkers in a real-life setting. Further, the implications of such transitions to remote health care on real-life clinical practice patterns and patient care and outcomes need to be determined. This step is important given the lack of established guidelines outlining best practice for remote care and the potential for unintended consequences like those created by the so-called digital divide, whereby specific groups of patients, namely those who are elderly, are from ethnic minority groups, and have a higher prevalence of comorbidities, are likely less able to use remote care through lack of access to the internet or technology literacy. Conversely, digital health may have the potential for greater inclusivity and helping to break down existing societal divides. Older patients, those from ethnic minority backgrounds, or those with more medical comorbidities may find remote monitoring for health care purposes more attractive because they have greater difficulty accessing in-person visits due to barriers to transportation or scheduling.[76] Thus, the potential effects that voice biomarkers, and digital health care biomarkers in general, integrated into care practice models have on patient outcomes should be studied in large prospective trials.

As part of these studies, a greater understanding of the mechanisms that link detectable changes in voice/speech with various pathologic processes should be pursued as

well. Health care professionals will be more likely to use a biomarker to help inform their clinical decisions if they believe that they have some appreciation for how it relates to the underlying pathophysiology of a disease, and in the relationship between voice/speech analysis and diseases such as CAD, this remains unclear. It may be that voice/speech analysis provides information on an individual's general well-being or it could relate more closely to their emotional or psychosocial state, which could be associated with the underlying disease of interest. These relationships need to be disentangled. Another perspective relates to issues with using clinical tools that rely on artificial intelligence, including to what extent we feel comfortable with acting on the output of a data-processing system that we do not intuitively understand. Lastly, privacy concerns should be considered because voice/speech data is sensitive and may be used to uncover a person's identity, ethnicity, and, after all, health status. For example, insurers could use these systems to analyze speech samples without consent to discriminate against customers, while workplaces could act similarly when selecting prospective employees. Appropriate regulation will be vital, as will measures that encrypt recordings and prevent privacy leakage to address some of these ethical concerns.

CONCLUSION

Digital biomarkers, aided by artificial intelligence, will be used increasingly in medicine going forward. In this review, we have outlined why voice biomarkers offer promise in remote health care monitoring both in clinical practice, by providing adjunctive information to aid health care professionals make decisions about the detection, progression, and treatment of disease, and in research by offering useful surrogate end points for clinical trials. However, for voice biomarkers to find an established role in medicine in the future, much work remains, including the systematic and rigorous evaluation of voice biomarkers in demographically and linguistically diverse populations using large prospective clinical trials, defining how to most appropriately integrate voice biomarkers into existing clinical and research work models, and cooperating with industry, regulatory agencies, insurance companies, and, most importantly, our patients to ensure that these biomarkers are used safely and effectively while prioritizing data protection and individual privacy.

POTENTIAL COMPETING INTERESTS

Dr Lilach Lerman has received grant support from AstraZeneca and has served as a consultant for AstraZeneca, Butterfly Biosciences, Curespec, Longeveron, and Beren Therapeutics. The other authors report no competing interests.

ACKNOWLEDGMENTS

We thank Joseph R. Duffy, PhD, for his insightful comments during the preparation of the submitted manuscript.

The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions: Dr Sara—Conceptualization, Methodology, Investigation, Writing/original draft, Writing/review and editing, Visualization; Dr Orbelo—Conceptualization, Methodology, Investigation, Writing/review and editing, Visualization; Dr Maor—Methodology, Investigation, Writing/review and editing, Visualization; Dr Lliach O. Lerman—Conceptualization, Project administration, Supervision, Writing/review and editing, Visualization; Dr Amir Lerman—Conceptualization, Project administration, Supervision, Writing/review and editing, Visualization.

Abbreviations and Acronyms: **AUC,** area under the curve; **CAD,** coronary artery disease; **COVID-19,** coronavirus disease 2019; **CVD,** cardiovascular disease; **MFCC,** Mel Frequency Cepstral Coefficient; **PTSD,** posttraumatic stress disorder; **SAAI,** speech active acoustic in-person; **SADI,** speech active dual (linguistic and acoustic) in-person; **VAAR,** vocal active acoustic remote

Correspondence: Address to Amir Lerman, MD, Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (lerman.amir@mayo.edu).

ORCID

Lilach O. Lerman: https://orcid.org/0000-0002-3271-3887

REFERENCES

1. Uscher-Pines L, Sousa J, Jones M, et al. Telehealth use among safety-net organizations in California during the COVID-19 pandemic. *JAMA*. 2021;325(11):1106-1107.
2. Dorsey ER, Topol EJ. State of telehealth. *N Engl J Med*. 2016; 375(2):154-161.
3. Hunt TL II, Hooten WM. The effects of COVID-19 on telemedicine could outlive the virus. *Mayo Clin Proc Innov Qual Outcomes*. 2020;4(5):583-585.
4. Contreras CM, Metzger GA, Beane JD, Dedhia PH, Ejaz A, Pawlik TM. Telemedicine: patient-provider clinical engagement during the COVID-19 pandemic and beyond. *J Gastrointest Surg*. 2020;24(7):1692-1697.
5. Margolis KL, Asche SE, Dehmer SP, et al. Long-term outcomes of the effects of home blood pressure telemonitoring and pharmacist management on blood pressure among adults with uncontrolled hypertension: follow-up of a cluster randomized clinical trial. *JAMA Netw Open*. 2018;1(5): e181617.
6. Merriel SWD, Andrews V, Salisbury C. Telehealth interventions for primary prevention of cardiovascular disease: a systematic review and meta-analysis [published correction appears in *Prev Med*. 2014;67:343]. *Prev Med*. 2014;64:88-95.
7. Ringeval F, Shuller B, Valstar M, et al. AVEC 2019 Workshop and Challenge: State-of-mind, detecting depression with AI, and cross-cultural affect recognition. In: *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*. Association for Computing Machinery; 2019:3-12.
8. Rutowski T, Harati A, Lu Y, Shriberg E. Optimizing speech-input length for speaker-independent depression classification. *Interspeech*. 2019:3023-3027.
9. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun*. 2015; 71:10-49.
10. Pampouchidou A, Simantiraki O, Fazlollahi A, et al. Depression assessment by fusing high and low level features from audio, video, and text. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. Association for Computing Machinery; 2016:27-34.
11. Williamson JR, Godoy E, Cha M, et al. Detecting depression using vocal, facial and semantic communication cues. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. Association for Computing Machinery; 2016:11-18.
12. Sanz L, Sistiaga JA, Lara AJ, Cuende E, García-Alcántara F, Rivera T. The prevalence of dysphonia, its association with immunomediated diseases and correlation with biochemical markers. *J Voice*. 2012;26(2):148-153.
13. Hamdan A-l, Jabbour J, Nassar J, Dahouk I, Azar ST. Vocal characteristics in patients with type 2 diabetes mellitus. *Eur Arch Otorhinolaryngol*. 2012;269(5):1489-1495.
14. Mohammadzadeh A, Heydari E, Azizi F. Speech impairment in primary hypothyroidism. *J Endocrinol Invest*. 2011;34(6): 431-433.
15. Murton OM, Hillman RE, Mehta DD, et al. Acoustic speech analysis of patients with decompensated heart failure: a pilot study. *J Acoust Soc Am*. 2017;142(4):EL401.
16. Orlikoff RF. The relationship of age and cardiovascular health to certain acoustic characteristics of male voices. *J Speech Hear Res*. 1990;33(3):450-457.
17. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med*. 2005;352(16):1685-1695.
18. Liao D, Cai J, Rosamond WD, et al. Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study; the ARIC Study. *Am J Epidemiol*. 1997;145(8): 696-706.
19. Tsuji H, Larson MG, Venditti FJ Jr, et al. Impact of reduced heart rate variability on risk for cardiac events: the Framingham Heart Study. *Circulation*. 1996;94(11):2850-2855.
20. Hackett RA, Steptoe A. Type 2 diabetes mellitus and psychological stress - a modifiable risk factor. *Nat Rev Endocrinol*. 2017;13(9):547-560.
21. Kivimäki M, Kawachi I. Work stress as a risk factor for cardiovascular disease. *Curr Cardiol Rep*. 2015;17(9):630.
22. Dragano N, Siegrist J, Nyberg ST, et al; IPD-Work Consortium. Effort-reward imbalance at work and incident coronary heart disease: a multicohort study of 90,164 individuals. *Epidemiology*. 2017;28(4):619-626.
23. Kivimäki M, Jokela M, Nyberg ST, et al; IPD-Work Consortium. Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. *Lancet*. 2015; 386(10005):1739-1746.
24. Huang Y, Xu S, Hua J, et al. Association between job strain and risk of incident stroke: a meta-analysis. *Neurology*. 2015;85(19): 1648-1654.
25. Johannes B, Wittels P, Enne R, et al. Non-linear function model of voice pitch dependency on physical and mental load. *Eur J Appl Physiol*. 2007;101(3):267-276.
26. Protopapas A, Lieberman P. Fundamental frequency of phonation and perceived emotional stress. *J Acoust Soc Am*. 1997; 101(4):2267-2277.
27. Holmqvist-Jämsén S, Johansson A, Santtila P, Westberg L, von der Pahlen B, Simberg S. Investigating the role of salivary cortisol on vocal symptoms. *J Speech Lang Hear Res*. 2017;60(10):2781-2791.
28. Vermani M, Marcus M, Katzman MA. Rates of detection of mood and anxiety disorders in primary care: a descriptive, cross-sectional study. *Prim Care Companion CNS Disord*. 2011; 13(2):PCC.10m01013.
29. Carey M, Jones K, Meadows G, et al. Accuracy of general practitioner unassisted detection of depression. *Aust N Z J Psychiatry*. 2014;48(6):571-578.
30. Sara JDS, Lerman LO, Lerman A. The endothelium is a key player in the vascular response to acute mental stress [editorial]. *Eur Heart J*. 2021;42(39):4089-4091.
31. Kivimäki M, Steptoe A. Effects of stress on the development and progression of cardiovascular disease. *Nat Rev Cardiol*. 2018;15(4):215-229.
32. Maor E, Sara JD, Orbelo DM, Lerman LO, Levanon Y, Lerman A. Voice signal characteristics are independently associated with coronary artery disease. *Mayo Clin Proc*. 2018; 93(7):840-847.

33. Sara JDS, Maor E, Borlaug B, et al. Non-invasive vocal biomarker is associated with pulmonary hypertension. *PLoS One*. 2020; 15(4):e0231441.
34. Sara JDS, Maor E, Orbelo D, Gulati R, Lerman LO, Lerman A. Noninvasive voice biomarker is associated with incident coronary artery disease events at follow-up. *Mayo Clin Proc*. 2022; 97(5):835-846.
35. Gerratt BR, Kreiman J, Garellek M. Comparing measures of voice quality from sustained phonation and continuous speech. *J Speech Lang Hear Res*. 2016;59(5):994-1001.
36. Maor E, Perry D, Mevorach D, et al. Vocal biomarker is associated with hospitalization and mortality among heart failure patients. *J Am Heart Assoc*. 2020;9(7):e013359.
37. Picone JW. Signal modeling techniques in speech recognition. *Proc IEEE*. 1993;81(9):1215-1247.
38. O'Shaughnessy DD. Invited paper: Automatic speech recognition: history, methods and challenges. *Pattern Recogn*. 2008; 41(10):2965-2979.
39. Godino-Llorente JI, Gómez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Trans Biomed Eng*. 2006;53(10): 1943-1953.
40. Eskidere Ö, Gürhanli A. Voice disorder classification based on multitaper Mel Frequency Cepstral Coefficients features. *Comput Math Methods Med*. 2015;2015:956249.
41. Trevisan J, Angelov PP, Carmichael PL, Scott AD, Martin FL. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst*. 2012; 137(14):3202-3215.
42. Dashtipour K, Tafreshi A, Lee J, Crawley B. Speech disorders in Parkinson's disease: pathophysiology, medical management and surgical approaches. *Neurodegener Dis Manag*. 2018;8(5):337-348.
43. Harel B, Cannizzaro M, Snyder PJ. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. *Brain Cogn*. 2004; 56(1):24-29.
44. Tracy JM, Özkanca Y, Atkins DC, Hosseini Ghomi R. Investigating voice as a biomarker: deep phenotyping methods for early detection of Parkinson's disease. *J Biomed Inform*. 2020; 104:103362.
45. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng*. 2012;59(5): 1264-1271.
46. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng*. 2010; 57(4):884-893.
47. Cushnie-Sparrow D, Adams S, Abeyesekera A, Pieterman M, Gilmore G, Jog M. Voice quality severity and responsiveness to levodopa in Parkinson's disease. *J Commun Disord*. 2018;76: 1-10.
48. Forbes-McKay KE, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurol Sci*. 2005;26(4):243-254.
49. Oulhaj A, Wilcock GK, Smith AD, de Jager CA. Predicting the time of conversion to MCI in the elderly: role of verbal expression and learning. *Neurology*. 2009;73(18): 1436-1442.
50. Reilly J, Peelle JE, Antonucci SM, Grossman M. Anomia as a marker of distinct semantic memory impairments in Alzheimer's disease and semantic dementia. *Neuropsychology*. 2011;25(4):413-426.
51. Ahmed S, Haigh A-MF, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*. 2013;136(pt 12):3727-3737.
52. Toth L, Hoffmann I, Gosztolya G, et al. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res*. 2018;15(2):130-138.
53. Martínez-Sánchez F, Meilán JJG, Carro J, Ivanova O. A prototype for the voice analysis diagnosis of Alzheimer's disease. *J Alzheimers Dis*. 2018;64(2):473-481.
54. König A, Satt A, Sorin A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimers Dement (Amst)*. 2015;1(1): 112-124.
55. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis*. 2016; 49(2):407-422.
56. Bonneh YS, Levanon Y, Dean-Pardo O, Lossos L, Adini Y. Abnormal speech spectrum and increased pitch variability in young autistic children. *Front Hum Neurosci*. 2011;4:237.
57. Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open J Eng Med Biol*. 2020;1:275-281.
58. Maor E, Tsur N, Barkai G, et al. Noninvasive vocal biomarker is associated with severe acute respiratory syndrome coronavirus 2 infection. *Mayo Clin Proc Innov Qual Outcomes*. 2021; 5(3):654-662.
59. Marmar CR, Brown AD, Qian M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. *Depress Anxiety*. 2019;36(7):607-616.
60. Belouali A, Gupta S, Sourirajan V, et al. Acoustic and language analysis of speech for suicidal ideation among US veterans. *BioData Min*. 2021;14(1):11.
61. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016.
62. Zhang L, Duvvuri R, Chandra KKL, Nguyen T, Ghomi RH. Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depress Anxiety*. 2020;37(7):657-669.
63. Taguchi T, Tachikawa H, Nemoto K, et al. Major depressive disorder discrimination using vocal acoustic features. *J Affect Disord*. 2018;225:214-220.
64. Yamamoto M, Takamiya A, Sawada K, et al. Using speech recognition technology to investigate the association between timing-related speech features and depression severity. *PLoS One*. 2020;15(9):e0238726.
65. Hashim NW, Wilkes M, Salomon R, Meggs J, France DJ. Evaluation of voice acoustics as predictors of clinical depression scores. *J Voice*. 2017;31(2):256.e1-256.e6.
66. Lee S, Suh SW, Kim T, et al. Screening major depressive disorder using vocal acoustic features in the elderly by sex. *J Affect Disord*. 2021;291:15-23.
67. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. *Biol Psychiatry*. 2012;72(7):580-587.
68. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics*. 2007;20(1):50-64.
69. Faurholt-Jepsen M, Busk J, Frost M, et al. Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry*. 2016;6(7):e856.

70. Weiner L, Guidi A, Doignon-Camus N, Giersch A, Bertschy G, Vanello N. Vocal features obtained through automated methods in verbal fluency tasks can aid the identification of mixed episodes in bipolar disorder. *Transl Psychiatry*. 2021; 11(1):415.
71. 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.
72. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit Med*. 2020;3:55.
73. Sharma G, Umapathy K, Krishnan S. Trends in audio signal feature extraction methods. *Appl Acoustics*. 2020;158:107020.
74. Sajal MSR, Ehsan MT, Vaidyanathan R, Wang S, Aziz T, Mamun KAA. Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis. *Brain Inform*. 2020;7(1):12.
75. Arora S, Baghai-Ravary L, Tsanas A. Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice. *J Acoust Soc Am*. 2019;145(5):2871.
76. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. *J Community Health*. 2013;38(5):976-993.