Digital Biomarkers

Research Report

Digit Biomark 2026;10:63–73
DOI: 10.1159/000550566

Received: October 6, 2025
Accepted: January 12, 2026
Published online: March 6, 2026

Voice Signatures of Momentary Psychological Stress in Real-Life Environments: Results from the Colive Voice Study

Noémie Topalian[a, b] Abir Elbeji[a] Charline Bour[a, c] Hanin Ayadi[a] Mégane Pizzimenti[a] Camille Perchoux[b] Guy Fagherazzi[a]

[a]Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health (LIH), Strassen, Luxembourg; [b]Urban Development and Mobility, Luxembourg Institute of Socio-Economic Research (LISER), Esch-sur-Alzette, Luxembourg; [c]Fonds de Dotation Clinatec, Université Grenoble Alpes, Grenoble, France

Keywords

Stress · Vocal biomarkers · Mental/psychological health

Abstract

Introduction: Voice is hypothesized to be modulated by stress and thus could be used as a potential stress detection and monitoring solution. In the literature, vocal biomarkers for stress have mostly been developed on experimental data, with limited samples. Therefore, this study aimed to present insights into the effect of momentary psychological stress on voice in real-life recordings, across different languages, genders, and vocal tasks. ***Methods:*** Participants from the Colive Voice study reported their stress level on a 1 to 5 Likert scale. Two tasks were performed: a text reading task and an A-vowel phonation. We analyzed the data cross-sectionally. We extracted vocal features with the DisVoice library and performed ordinary least squares regression models to evaluate the association of vocal features with stress. Models were stratified by gender and language (French/English) and controlled for age, smoking status, alcohol consumption, the presence of chronic disease, education level, mother tongue, well-being, fatigue, and depression. Benjamini-Hochberg correction was applied to control for multiple testing. ***Results:*** We analyzed a sample of 4,155 participants, 2,011 in French (1,621 women, 390 men) and 2,144 in English (1,105 women, 1,039 men). In the text reading task, we found that stress was associated with two articulatory features for English-speaking women. Among French-speaking women, higher stress was linked with lower pitch and higher shimmer. The duration of pauses and one glottal feature were also associated with stress. In the A-vowel phonation task, pitch and the variability of the pitch perturbation quotient were lower with stress in English-speaking men. French-speaking women had increased voice intensity and loudness with stress. ***Conclusion:*** We were able to confirm the association of momentary psychological stress with various vocal features in real-life settings, but not across languages, vocal tasks, or gender. Future research should include longitudinal studies to investigate the potential of using voice as an intraindividual monitoring biomarker for stress.

Introduction

Stress is considered one of the main public health challenges of the century. While acute stress responses are a common and natural answer to external stimuli, the accumulation of stress – namely, chronic stress – is associated with several disease outcomes [1, 2].

karger@karger.com
www.karger.com/dib

Karger
OPEN ACCESS

Correspondence to:
Guy Fagherazzi, Guy.Fagherazzi@lih.lu

Vocal biomarkers are “a feature or combination of features in the voice that has been identified and validated as associated with a clinical outcome” [3]. The production of voice is affected by muscle tension, heart rate, and lung capacity, among others. Thus, the use of voice as a stress biomarker has gained consequent interest and several studies have shown the effects of stress on voice [4–6].

In the literature, we commonly find that an increase in the fundamental frequency F0 (pitch) is a marker of stress. Jitter and shimmer are also frequently studied as stability indicators of the fundamental frequency [4, 6]. Other speech markers, such as speed or the length of voiced segments, are also found in the literature [7–9].

Several studies developed prediction models for stress, using vocal features [10–13]. They mostly rely on experimental designs comparing non-stressed and stressed voices, using vocal features before and after an acute stressor. However, these studies are limited in the detection of stress in real-life settings.

Our analysis draws on data from the Colive Voice study [14], an international study for the diagnosis and monitoring of symptoms and diseases through voice. We aimed to investigate a wide range of vocal features and their association with self-reported momentary psychological stress in real-life recordings.

Methods

Study Population

We included participants from the Colive Voice study, a worldwide research project on vocal biomarkers for chronic pathologies and health symptoms initiated by the Luxembourg Institute of Health in 2021. Participants were enrolled between January 2022 and January 2024.

Colive Voice is an observational, cross-sectional study where participants answer in English, French, Spanish, German, Portuguese, or Arabic. Due to the limited number of participants in other languages, we solely performed our analysis on French and English speakers.

Eligible participants were above the age of 15, with and without health conditions. Participants were volunteers and recruited through online communication. They completed health and demographic questionnaires at home, on a smartphone, tablet, or computer and performed standardized vocal tasks in French or English. We included all participants in this analysis, with the exception of participants who were currently hospitalized for COVID.

Stress Measurement

Stress was assessed with the Single-Item Stress Question (SISQ): “Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because his/her mind is troubled all the time. Do you feel this kind of stress these days?” and was measured on a Likert scale as “Not at all,” “Only a little,” “To some extent,” “Rather much,” and “Very much” [15]. We recoded these labels as a score ranging from 1 to 5. This single item has been shown to be associated with the risk of future depression, sickness, or injuries [16].

Covariates

In the health and demographics questionnaires, participants provided their age, gender, and preferred language for the study. We stratified the population by language and gender (English/French-speaking and women/men). Several variables, hypothesized to influence both stress and voice production, were also collected in the questionnaire. Thus, we adjusted our models on the following confounding factors: age, education level, smoking, alcohol consumption, the presence of chronic disease, mother tongue, well-being, depression, and fatigue [3, 17–19].

Education level was defined as the highest level of education completed. We divided it into two categories: high school (“some high school or lower,” “high school”) and university education (“bachelor’s degree,” “master’s degree,” “PhD or higher”). Smoking was defined as current (daily or occasional) or nonsmoker (including past smokers). Alcohol was a numerical variable, computed as the number of glasses of alcohol per week. Chronic disease was defined as the diagnosis of several or any of the following diseases: cancer, respiratory disorder (emphysema, chronic bronchitis, chronic obstructive pulmonary disease, asthma), cardiovascular disorder (infarction/heart attack, congestive heart failure, coronary heart disease, angina pectoris, hypertension), endocrine disorder (diabetes, thyroidic disease, chronic kidney disease), neurological disorder (rheumatoid arthritis or systemic lupus, multiple sclerosis, epilepsy, amyotrophic lat. sclerosis, narcolepsy, Parkinson’s disease, stroke, migraine), gastrointestinal disease (hepatitis, Crohn’s disease, ulcerative colitis), psychological disorder (depression), or any other disorder affecting voice (laryngitis, noncancerous vocal cord lesion, laryngopharyngeal reflux, leukoplakia). Mother tongue was reported by the participants and defined as “native” if the questionnaire was done in the same language, and “non-native,” otherwise. Well-being was measured using the 5-

item questionnaire WHO-5, with scores ranging from 0 to 25, 25 being the best well-being possible [20]. Depression was assessed using the PHQ-9 questionnaire, with a score from 0 to 27, and higher scores indicate more severe depression [21]. Fatigue was measured using the Fatigue Severity Scale (FSS), with a score ranging from 9 to 63, with higher scores indicating a more severe fatigue [22].

Vocal Tasks

Participants were asked to place themselves in a quiet environment and to complete several short standardized vocal tasks. The vocal tasks were performed directly following the health questionnaires and could not be delayed. In this analysis, we used the text reading task (extract from Article 25 of the Human Rights Declaration [23]) and an A-vowel phonation task. The text was chosen for its neutrality of tone and availability in several languages. This task has been shown to perform well to predict fatigue or type 2 diabetes [24, 25].

- Text reading: Press "Record" and read the following passage. Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age, or other lack of livelihood in circumstances beyond his control.
- A-vowel phonation: Now take a deep breath, press "Record," then say the sound "Aaaah" for as long as possible. Stop recording when you run out of breath.

Vocal Feature Extraction

Raw audio files were preprocessed and quality-checked to ensure consistency and harmonization of the recordings. Vocal features were extracted using the DisVoice library in Python 3.8.20, version 0.1.8. We extracted all the features from the Prosody, Phonation, Articulation, Phonological, and Glottal sets. To ensure explainability of the features, we limited the functionals to mean and standard deviation (SD). Due to recording quality, limited voiced material, or algorithmic constraints in DisVoice, some features failed during extraction. Participants with more than 75% of missing or zero values across all feature sets for any of their audio recordings were excluded from the analysis.

To reduce redundancy among features, we removed those that were highly correlated. We iteratively identified groups of features that were highly correlated (Spearman's coefficient >0.9) from a correlation matrix. Within each group, we iteratively removed the features with the highest variance inflating factor until all features had a variance inflating factor <5.

For the A-vowel phonation task, we left out articulation and phonological features, as they are only relevant in continuous speech [26]. The total number of vocal features for each vocal task is available in Table 1.

Data Analysis

We described the population's characteristics per language and gender. We used mean (SD) for numerical variables and number (%) for categorical variables. We tested for significant differences between stress categories. We performed the Kruskal-Wallis test and chi-square test for numerical and categorical variables, respectively.

All the numerical variables and vocal features were standardized using a standard scaler before the analysis. We identified confounding variables and fitted our models as follows:

- M0: Vocal feature ~ stress
- M1: Vocal feature ~ stress + age + education level + smoking + alcohol consumption + chronic disease + mother tongue
- M2: Vocal feature ~ stress + age + education level + smoking + alcohol consumption + chronic disease + mother tongue + well-being score + depression score + fatigue severity score.

We performed ordinary least squares regressions over each vocal feature to evaluate their association with stress, using the statsmodels library version 0.14.0. The reference categories for categorical variables were non-smoker, high school education, no chronic disease, and native speaker. We reported the regression coefficients, confidence intervals, and *p* values for each model. The *p* values were adjusted for multiple testing, using the Benjamini-Hochberg method.

Results

Study Population

In January 2024, $n = 5{,}379$ participants were included in the Colive Voice study. We excluded $n = 27$ currently hospitalized for COVID, and $n = 1{,}197$ were excluded for insufficient audio quality or missing values in the feature extraction step. In total, we analyzed a sample of

Table 1. Number of nonredundant vocal features analyzed by vocal task

Text reading task	prosody (43 features)	phonation (14 features)	articulation (244 features)	phonological (36 features)	glottal (18 features)
Women					
English (n = 1,105)	11	3	66	9	3
French (n = 1,621)	11	3	65	9	2
Men					
English (n = 1,039)	11	3	68	9	4
French (n = 390)	11	3	66	9	4

A-vowel phonation	prosody (43 features)	phonation (14 features)	glottal (18 features)
Women			
English (n = 1,105)	14	5	6
French (n = 1,621)	16	5	7
Men			
English (n = 1,039)	19	5	6
French (n = 390)	19	5	5

4,155 participants, 2,011 in French (1,621 women, 390 men) and 2,144 in English (1,105 women, 1,039 men). The participant inclusion flowchart is available in Figure 1.

The population had received in majority higher education (68%), and the prevalence of at least one chronic disease was high (78%). The population description is available in Table 2.

In both languages, women systematically reported higher stress levels compared to men. Figure 2 shows the proportion of stress levels by population. The distribution of stress was more right-skewed for women and for English-speaking participants overall, with more participants reporting moderate-to-high stress in those populations.

Association of Vocal Features with Stress

Figure 3 shows the regression coefficients, confidence intervals, and adjusted p values of the stress scores for each regression over a vocal feature. We found no common features across gender, language, or vocal tasks. For interpretability, we computed the mean and SD of each feature, available in online supplementary Tables S1 and S2 (for all online suppl. material, see https://doi.org/10.1159/000550566), for each vocal task. In the following results, we converted standardized effects into raw effects. We also present the results of the progressive adjustments of models M0 and M1 in online supplementary Tables S3 and S4 and the standardized effects in online supplementary Figure S1.

Text Reading Task

In the text reading task, we found no associations between stress and vocal features for men, in any language.

Women, English-speaking. English-speaking women had 2 articulation features significantly associated with stress: average Delta Delta Formant 1 (DDF1) and average Mel-frequency cepstral coefficient (MFCC) 7. In raw units, a 1-point increase in stress was associated with a 0.05 (0.02, 0.07) Hz/frame^2 increase in average DDF1. This meant the acceleration in Formant 1 was higher with each increase in stress, suggesting a high variability in F1. DDF1 is computed over each frame, and the unit is in Hz per frame^2, with each frame being 40 ms, using DisVoice articulation default parameters [26].

MFCCs represent the power spectrum of a sound, using a scale imitating how humans perceive pitch and loudness. MFCC 7 is a middle-to-high order of coefficient that captures rapid variations in the spectral envelope across frequency. The average MFCC 7 was 0.28 (0.12, 0.43) times higher with each 1-point increase in stress.

Women, French-speaking. In French-speaking women, we found 1 prosody, 1 phonation, 1 phonological, and 1

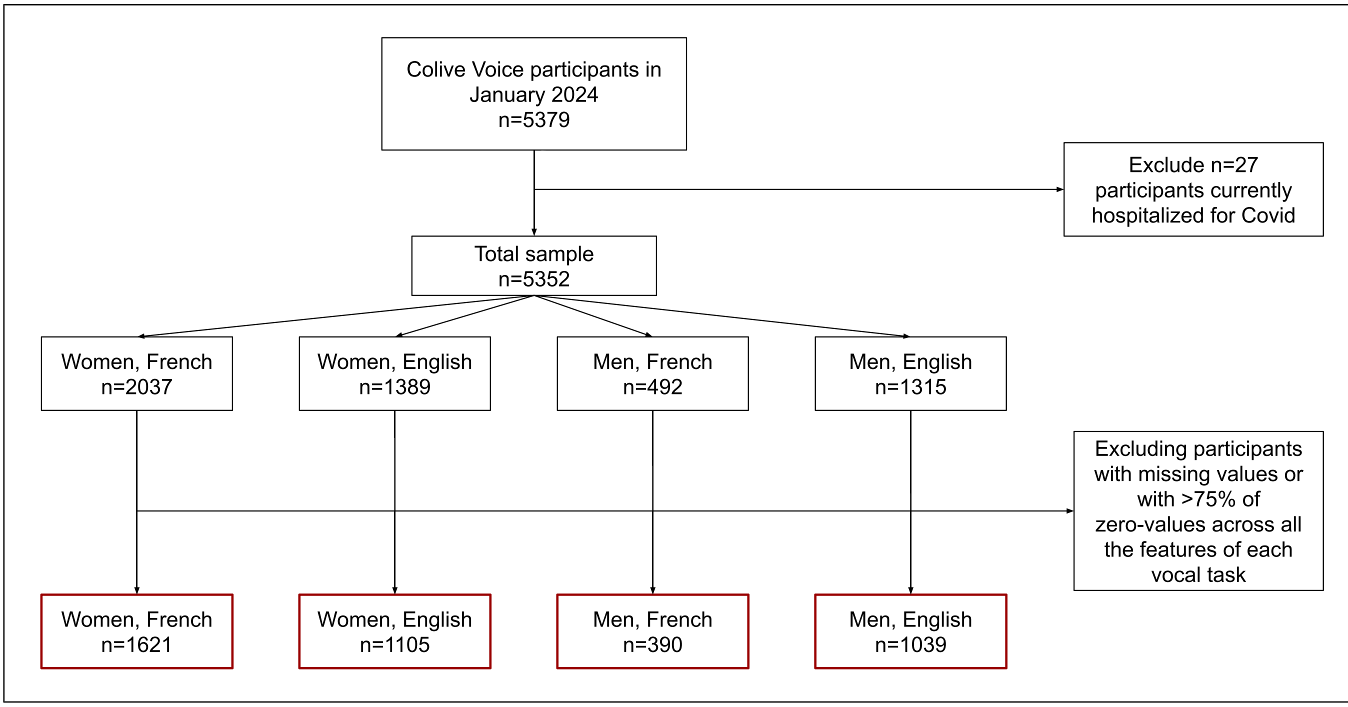

Fig. 1. Inclusion flowchart of the Colive Voice participants included in the analysis.

Table 2. Description (mean [SD], number [%]) of Colive Voice participants included in the analysis

	Total sample	Women		Men		*p* value
		English	French	English	French	
Participants, *N*	4,155	1,105	1,621	1,039	390	–
Stress score (mean [SD])	2.6 (1.2)	2.8 (1.1)	2.7 (1.2)	2.4 (1.1)	2.1 (1.1)	<0.001
Age, years (mean [SD])	44.5 (15.3)	43.6 (14.5)	44.3 (15.1)	44.0 (14.8)	48.9 (18.3)	<0.001
Higher education, *n* (%)	2,839 (68.3)	681 (61.6)	1,239 (76.4)	659 (63.4)	260 (66.7)	<0.001
Alcohol consumption, glasses per week (mean [SD])	3.7 (11.6)	2.8 (6.1)	3.0 (13.1)	5.6 (13.3)	4.2 (11.1)	<0.001
Current smoker, *n* (%)	710 (17.1)	186 (16.8)	243 (15.0)	231 (22.2)	50 (12.8)	<0.001
Chronic disease, *n* (%)	3,258 (78.4)	858 (77.6)	1,366 (84.3)	703 (67.7)	331 (84.9)	<0.001
Non-native speaker, *n* (%)	384 (9.2)	102 (9.2)	126 (7.8)	85 (8.2)	71 (18.2)	<0.001
Well-being score (WHO-5) (mean [SD])	12.3 (5.7)	12.3 (5.7)	11.4 (5.4)	13.4 (6.0)	12.9 (5.6)	<0.001
Fatigue score (FSS) (mean [SD])	35.9 (14.0)	34.6 (13.4)	40.0 (14.2)	31.8 (12.6)	34.1 (14.0)	<0.001
Depression score (PHQ-9) (mean [SD])	7.7 (6.0)	7.3 (6.1)	9.1 (5.9)	6.2 (6.0)	6.9 (5.3)	<0.001

Stress was measured on a 1–5 Likert scale. Higher education is defined as completing a bachelor's, master's, or PhD degree. Chronic disease was defined as having had any cancer, pulmonary, endocrine, cardiovascular, neurological, gastrointestinal, mental disorder, or any disorder affecting voice. All the variables were self-reported. The *p* values indicate differences between the 4 populations, computed with the Kruskal-Wallis test and chi-square test for numerical and categorical variables, respectively. FSS, Fatigue Severity Scale.

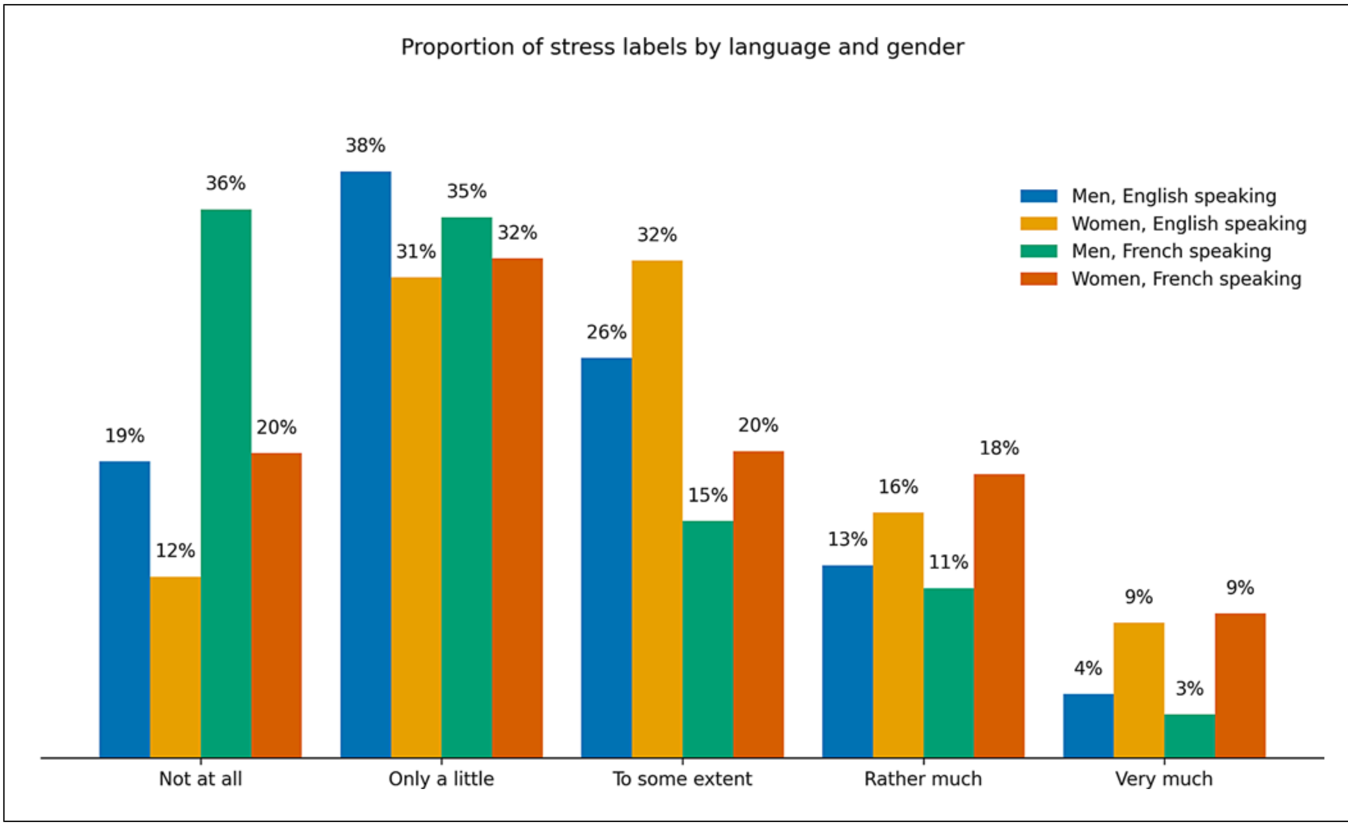

Fig. 2. Histogram of stress score distribution, by gender and language spoken.

glottal features associated with stress. The average F0 tilt is a linear estimation of the fundamental frequency (i.e., the trend of the pitch) over each frame of the recording. It was lower with each 1-point stress increase (−4.64 [−7.74, −1.55]).

Shimmer describes the variations in the amplitude of the voice signal, i.e., the loudness of the signal from one glottal cycle to the next [27]. The average shimmer was positively associated with stress (0.15 [0.05, 0.25]).

The mean duration of pauses (silences in speech) decreased with stress (−0.05 [−0.08, −0.02]).

Glottal Closure Instants (GCIs) indicate the time points when vocal folds close completely, tracking their vibration cycles. The global average variability between GCI was higher with stress (0.00007 [0.00002, 0.00012]).

A-Vowel Phonation

In the A-vowel phonation task, we found no associations between stress and vocal features for English-speaking women and French-speaking men.

Men, English-speaking. In English-speaking men, mean F0 contour, the overall pitch of the recording, was negatively associated with stress (−4.25 [−6.87, −1.64]). Pitch perturbation quotient (PPQ) is a measure of jitter, the stability in pitch, over several glottal cycles [27]. The SD of PPQ was also lower with stress (−0.28 [−0.41, −0.14]).

Women, French-speaking. Among French-speaking women, one phonation feature was significantly associated with stress: the average logarithmic energy (0.53 [0.18, 0.88]). Logarithmic energy captures overall loudness and vocal intensity.

Discussion

Our results showed that interindividual differences exist in the association between stress and voice, but that these associations are dependent on gender, language, and vocal task. Stress influences voice through different pathways: neurological, respiratory, endocrine, and cognitive. The stress response happens in the autonomic nervous system, which includes the sympathetic and parasympathetic branches. The vagus nerve, a key

Digit Biomark 2026;10:63–73
DOI: 10.1159/000550566

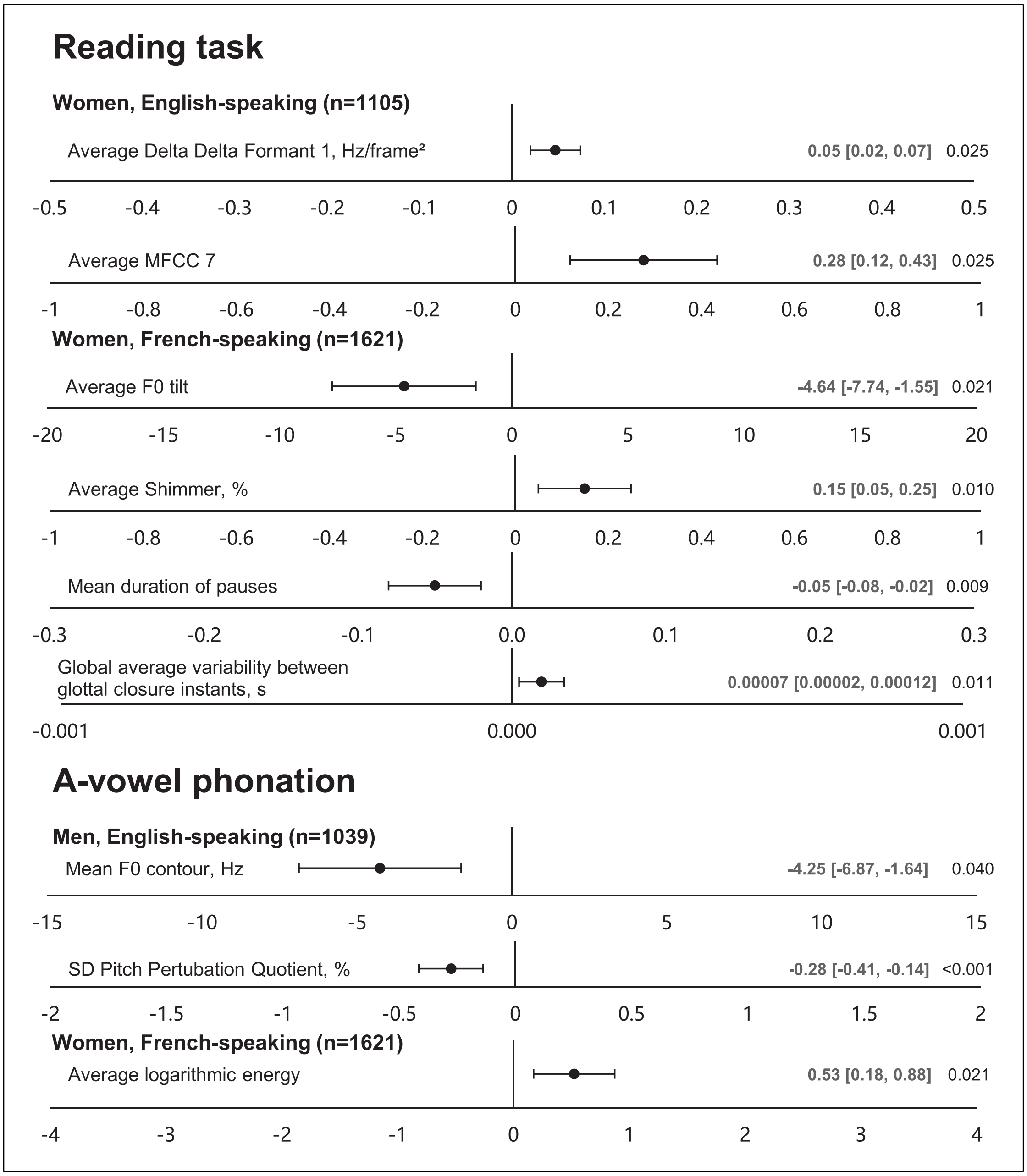

Fig. 3. Associations of vocal features with stress. Ordinary least squares regression coefficients (95% CIs, p values) for the association of stress with each vocal feature. There were n = 1,105 English-speaking women, n = 1,621 French-speaking women, and n = 1,039 English-speaking men. The models were adjusted for age, education level, alcohol consumption, smoking, chronic disease, mother tongue, fatigue, depression, and well-being score (M2 model). A one-unit increase in stress indicates a one-unit change in the vocal feature given by the regression coefficient.

component of the parasympathetic system, helps regulate these processes. It has been shown that stimulation of the vagus nerve affects voice production by affecting laryngeal muscle tension and vocal folds [28]. Muscular tension affecting voice, known as muscle tension dysphonia, can result from excessive activation or dysregulation of the laryngeal muscles, often exacerbated by stress [5].

Voice changes due to stress also happen in relation to the endocrine system. Changes in cortisol and other stress hormone levels have been linked to changes in vocal features [4]. Additionally, the gut-brain axis is hypothesized to influence stress responses through the activation of the hypothalamic-pituitary-adrenal axis, resulting in muscle tension or respiratory stability affecting voice [29]. Gender differences in stress-related voice changes may be influenced by women's hormonal cycles, in how stress is reported, and in the laryngeal system's reactivity to vocal strain [30].

Delta Delta Formant 1

We found a positive association between stress and average DDF1. This suggests quicker acceleration (changes in the rate of change) of Formant 1. High values indicate sharp vowel transitions, emotional speech, or quicker articulatory movements [31]. Results from the literature suggest that F1 increases under a stressful condition [32], as well as formant spacing [7]. However, we found no studies investigating DDF1 specifically.

Mel-Frequency Cepstral Coefficient 7

Our results showed that higher stress was associated with an increase in average MFCC 7. MFCC 7 captures spectral variability in a middle to high range, especially sensitive to consonant articulation, noise, and emotional distinctions [31, 33]. The physiological interpretation of these findings is limited due to the abstract nature of MFCCs. MFCCs have been used to train prediction models of stress using vocal features [12, 34]. In two studies, adding MFCCs to the feature set increased the performance of the model [35, 36]. In a study using a sample of voice actors undergoing a stressful task, Arushi and Teoh [10] found that MFCCs were the features with the most differences between the "confident" and "stressed" vocal expressions.

Average F0 Tilt

We found a negative association between stress and the average tilt of a linear estimation of F0. It captures speech direction and expressiveness [31, 37]. This indicates that with higher stress, F0 is flatter or decreases, which would typically be more associated with fatigue or sadness. F0 is the most commonly studied feature in the analysis of stress and voice. It has been used in prediction models [11, 35]. However, the association between F0 and voice tends to be positive [4, 6]. Several studies found an increase in F0 under stressful conditions [7, 9, 32, 38–43]. Limited literature exists on the between-person effects of stress on voice. While F0 also increases with stress between participants, one study showed that at baseline, F0 is not associated with cortisol and that this positive association becomes true once baseline cortisol is doubled [43]. A longitudinal study found no associations between stress and F0, whether within or between participants [8].

Shimmer

On average, shimmer increased with stress. High shimmer indicates irregularities in vocal intensity. Some studies found no change in shimmer in the stress condition [9, 32]. One study found that shimmer was lower in a negative-feedback condition [39]. A longitudinal study showed that shimmer was positively associated with long-term stressors at the individual level but not with perceived stress or work stressors [8].

Pause Duration

We found that the mean pause duration was lower with stress. This feature describes silence periods in speech [44]. One study found contradictory results where pause time was higher in the stressful condition [45]. Similarly to F0, this difference could be due to the modeling of intra- versus interindividual effects.

Global Variability of GCIs

The global variability of GCI increased with stress. This indicates irregular vocal fold vibration, which may be linked to stress, fatigue, or voice disorders [46]. This feature reflects vocal fold consistency but has not been studied in the case of stress.

Mean F0 Contour

The mean F0 contour describes the overall pitch of the voice recording. We found that it was negatively associated with stress. Similarly to F0 tilt, this finding is contradictory with most of the literature using stress-inducing experiments [7, 9, 32, 38–43]. This difference could be attributed to the between-participant comparisons or the assessment of stress, which captures more chronic aspects of stress instead of an acute stressor [15].

Pitch Perturbation Quotient

We found that the SD of PPQ decreased with stress. PPQ is a measure of jitter over several glottal cycles. The SD of PPQ being lower suggests a more controlled and consistent voice. Jitter was used in prediction models [35]. A review identified a reduction in jitter as commonly reported with stress [4], along with a more recent study [9]. One longitudinal study found no association between stress and jitter [8].

Logarithmic Energy

The average logarithmic energy was positively associated with stress. One study found that within participants, elevated voice intensity was associated with increased work stressors [8].

Speech Rate

We found no association between stress and the number of words. One study showed that words-per-minute was higher in the stress condition [7]. Speech rate, defined as the number of syllables per second, and voiced segments per second, a proxy for speech rate, also increased with stressors [8, 9]. On the contrary, two studies found that the communication rate did not change between baseline and stress conditions [39, 45].

Nonfluencies and Linguistic Complexity

Nonfluencies (hesitations in speech) and linguistic complexity were not investigated in our study. The literature shows that nonfluencies are higher in the stress condition, and that stress is linked to lowest linguistic complexity in speech [45, 47].

Harmonics-To-Noise Ratio

We found no association between harmonics-to-noise ratio and stress, which aligns with a previous study [9], while another found that HRN was higher in the stress condition [39].

Mean Voiced Length

The length of voiced segments was not associated with stress. Kappen et al. [39] found similar results, while Kappen et al. [9] showed that mean voiced length increased with stress.

Strengths and Limitations

Our study allowed us to investigate various vocal features and their association with stress, in a large international study. This cross-sectional analysis allowed us to look at interindividual effects of stress on voice, in real-life settings, and with a large variety of recording devices, which enhances the ecological validity of our findings. The association of features was controlled for confounding factors (age, smoking status, alcohol consumption, the presence of chronic disease, mother tongue, education level, well-being, fatigue, and depression) and corrected for multiple testing to avoid the likelihood of results being attributed to data dredging.

We did not find common vocal biomarkers of stress between languages, vocal tasks, or genders. While small, the strength of the associations may still be relevant when considered alongside other acoustic markers of stress.

Stress was assessed using a single item and could be subject to a recall bias due to its phrasing, but it has been shown to be a reliable estimate of momentary and daily stress, and most correlated with work-related stress [15]. The order of the vocal tasks was not randomized. However, the risks of habituation and vocal strain are limited, respectively, due to the distinct nature of each task, and their length (less than 30 s).

While we were able to control our analysis for mother tongue, additional biases such as cultural and regional differences may still affect voice. The neutral and guided nature of the tasks partially mitigates this bias, but these differences should be investigated in further studies.

Conclusions and Perspectives for Future Work

This study highlighted the complexity of identifying vocal biomarkers for stress. While the fundamental frequency is commonly cited in the literature as associated with stress, most studies use stress-induction experiments and trigger acute stress. However, different stressors could trigger physiological reactions with varied intensity, influencing voice differently.

Longitudinal studies are now needed to further confirm the association of these vocal features with stress and disentangle the intra- and interindividual effects of stress on voice. Studies should have larger sample sizes to fully understand the linguistic, gender, and task-specific differences in the association of stress with vocal features.

Statement of Ethics

Colive Voice has been approved by the National Research Ethics Committee in Luxembourg (N°202103/01) in March 2021. Written informed consent was obtained electronically from all participants in the study. The Colive Voice study protocol is also registered in ClinicalTrials.gov (NCT04848623).

Conflict of Interest Statement

The authors have no conflict of interest to declare.

Funding Sources

N.T. and C.P. are funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This work is supported by ERC grant (ERC-2021-STG, FragMent, 101040492).

Author Contributions

G.F. contributed to the conception and design of the study. M.P. coordinated the data collection. H.A. and A.E. contributed to the preprocessing of the data. G.F., N.T., and C.P. contributed to the methodology. N.T., A.E. and C.B. contributed to the analysis of the data. N.T. drafted the article. G.F., M.P., A.E., and C.B. revised the manuscript critically. All authors read and approved the final manuscript.

Data Availability Statement

The data that support the findings of this study are not publicly available due to information that could compromise the privacy of research participants but are available from GF upon request. The code used to generate all the results is publicly available in the GitHub repository: https://github.com/LIHVOICE/Colive-Voice-Stress.git.

References

1 Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA. 2007;298(14):1685–7. https://doi.org/10.1001/jama.298.14.1685
2 Turner AI, Smyth N, Hall SJ, Torres SJ, Hussein M, Jayasinghe SU, et al. Psychological stress reactivity and future health and disease outcomes: a systematic review of prospective evidence. Psychoneuroendocrinology. 2020;114:104599. https://doi.org/10.1016/j.psyneuen.2020.104599
3 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. https://doi.org/10.1159/000515346
4 Giddens CL, Barron KW, Byrd-Craven J, Clark KF, Winter AS. Vocal indices of stress: a review. J Voice. 2013;27(3):390.E21–390.E29. https://doi.org/10.1016/j.jvoice.2012.12.010
5 Van Puyvelde M, Neyt X, McGlone F, Pattyn N. Voice stress analysis: a new framework for voice and effort in human performance. Front Psychol. 2018;9:1994. https://doi.org/10.3389/fpsyg.2018.01994
6 Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M. Review on psychological stress detection using biosignals. IEEE Trans Affect Comput. 2022;13(1):440–60. https://doi.org/10.1109/taffc.2019.2927337
7 Pisanski K, Sorokowski P. Human stress detection: cortisol levels in stressed speakers predict voice-based judgments of stress. Perception. 2021;50(1):80–7. https://doi.org/10.1177/0301006620978378
8 Langer M, König CJ, Siegel R, Fredenhagen T, Schunck AG, Hähne V, et al. Vocal-stress diary: a longitudinal investigation of the association of everyday work stressors and human voice features. Psychol Sci. 2022;33(7):1027–39. https://doi.org/10.1177/09567976211068110
9 Kappen M, Vanhollebeke G, Van Der Donckt J, Van Hoecke S, Vanderhasselt M-A. Acoustic and prosodic speech features reflect physiological stress but not isolated negative affect: a multi-paradigm study on psychosocial stressors. Sci Rep. 2024;14(1):5515. https://doi.org/10.1038/s41598-024-55550-3
10 Arushi DR, Teoh AN. Real-time stress detection model and voice analysis: an integrated VR-based game for training public speaking skills. 2021 IEEE conference on games (CoG). IEEE; 2021. https://doi.org/10.1109/cog52621.2021.9618989
11 Baird A, Triantafyllopoulos A, Zänkert S, Ottl S, Christ L, Stappen L, et al. An evaluation of speech-based recognition of emotional and physiological markers of stress. Front Comput Sci. 2021;3:750284. https://doi.org/10.3389/fcomp.2021.750284
12 Namkung J, Kim SM, Cho WI, Yoo SY, Min B, Lee SY, et al. Novel deep learning-based vocal biomarkers for stress detection in Koreans. Psychiatry Investig. 2024;21(11):1228–37. https://doi.org/10.30773/pi.2024.0131
13 Soury M, Devillers L. Stress detection from audio on multiple window analysis size in a public speaking task. 2013 Humaine Association conference on affective computing and intelligent interaction. IEEE; 2013. https://doi.org/10.1109/acii.2013.93
14 Home. Colive voice [Internet]; 2022. [cited 28 Jul 2025]. Available from: https://www.colivevoice.org/en/
15 Arapovic-Johansson B, Wåhlin C, Kwak L, Björklund C, Jensen I. Work-related stress assessed by a text message single-item stress question. Occup Med. 2017;67(8):601–8. https://doi.org/10.1093/occmed/kqx111
16 Salminen S, Kouvonen A, Koskinen A, Joensuu M, Väänänen A. Is a single item stress measure independently associated with subsequent severe injury: a prospective cohort study of 16,385 forest industry employees. BMC Public Health. 2014;14:543. https://doi.org/10.1186/1471-2458-14-543
17 Taylor S, Dromey C, Nissen SL, Tanner K, Eggett D, Corbin-Lewis K. Age-related changes in speech and voice: spectral and cepstral measures. J Speech Lang Hear Res. 2020;63(3):647–60. https://doi.org/10.1044/2019_JSLHR-19-00028
18 Ayadi H, Elbéji A, Despotovic V, Fagherazzi G. Digital vocal biomarker of smoking status using ecological audio recordings: results from the Colive Voice study. Digit Biomark. 2024;8(1):159–70. https://doi.org/10.1159/000540327
19 Almaghrabi SA, Clark SR, Baumert M. Bio-acoustic features of depression: a review. Biomed Signal Process Control. 2023;85:105020. https://doi.org/10.1016/j.bspc.2023.105020
20 Region hovedstaden. [cited 28 Jul 2025]. Available from: https://www.psykiatri-regionh.dk/who-5/who-5-questionnaires/Pages/default.aspx
21 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
22 Fatigue severity scale. In: Shirley Ryan AbilityLab [Internet]. [cited 28 Jul 2025]. Available from: https://www.sralab.org/rehabilitation-measures/fatigue-severity-scale
23 Website. Available from: https://www.un.org/en/about-us/universal-declaration-of-human-rights
24 Elbéji A, Zhang L, Higa E, Fischer A, Despotovic V, Nazarov PV, et al. Vocal biomarker predicts fatigue in people with COVID-19: results from the prospective Predi-COVID cohort study. BMJ Open. 2022;12(11):e062463. https://doi.org/10.1136/bmjopen-2022-062463
25 Elbéji A, Pizzimenti M, Aguayo G, Fischer A, Ayadi H, Mauvais-Jarvis F, et al. A voice-based algorithm can predict type 2 diabetes status in USA adults: findings from the Colive Voice study. PLOS Digit Health. 2024;3(12):e0000679. https://doi.org/10.1371/journal.pdig.0000679

26 Welcome to Disvoice's documentation! — DisVoice 0.1.9 documentation. [cited 28 Jul 2025]. Available from: https://disvoice.readthedocs.io/en/latest/index.html
27 Arias-Vergara T, Vásquez-Correa JC, Orozco-Arroyave JR. Parkinson's disease and aging: analysis of their effect in phonation and articulation of speech. Cognit Comput. 2017;9(6):731–48. https://doi.org/10.1007/s12559-017-9497-x
28 Charous SJ, Kempster G, Manders E, Ristanovic R. The effect of vagal nerve stimulation on voice. Laryngoscope. 2001;111(11 Pt 1):2028–31. https://doi.org/10.1097/00005537-200111000-00030
29 Foster JA, Rinaman L, Cryan JF. Stress & the gut-brain axis: regulation by the microbiome. Neurobiol Stress. 2017;7:124–36. https://doi.org/10.1016/j.ynstr.2017.03.001
30 Hunter EJ, Tanner K, Smith ME. Gender differences affecting vocal health of women in vocally demanding careers. Logoped Phoniatr Vocol. 2011;36(3):128–36. https://doi.org/10.3109/14015439.2011.587447
31 Vásquez-Correa JC, Orozco-Arroyave JR, Bocklet T, Nöth E. Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. J Commun Disord. 2018;76:21–36. https://doi.org/10.1016/j.jcomdis.2018.08.002
32 Sondhi S, Vijay R, Khan M, Salhan AK. Voice analysis for detection of deception. 2016 11th international conference on knowledge, information and creativity support systems (KICSS). IEEE; 2016; p. 1–6.
33 Orozco-Arroyave JR, Vásquez-Correa JC, Vargas-Bonilla JF, Arora R, Dehak N, Nidadavolu PS, et al. NeuroSpeech: an open-source software for Parkinson's speech analysis. Digit Signal Process. 2018;77:207–21. https://doi.org/10.1016/j.dsp.2017.07.004
34 Sharada A, Mamatha R, Meghana K, Monika A. Stress detection in women using speech analysis. Advances in engineering research. Dordrecht: Atlantis Press International BV; 2023; p. 797–808.
35 Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Campbell AT, et al. StressSense. Proceedings of the 2012 ACM Conference on Ubiquitous Computing. New York, NY, USA: ACM; 2012; p. 351–60.
36 Tomba K, Dumoulin J, Mugellini E, Abou Khaled O, Hawila S. Stress detection through speech analysis. Proceedings of the 15th international joint conference on e-Business and telecommunications. SCITEPRESS - Science and Technology Publications; 2018. https://doi.org/10.5220/0006855803940398
37 Dehak N, Dumouchel P, Kenny P. Modeling prosodic features with joint factor analysis for speaker verification. IEEE Trans Audio Speech Lang Process. 2007;15(7):2095–103. https://doi.org/10.1109/tasl.2007.902758
38 Finke JB, Zhang X, Plein D, Schilling TM, Schächinger H, Larra MF. Combining mental and physical stress: synergy or interference? Physiol Behav. 2021;233:113365. https://doi.org/10.1016/j.physbeh.2021.113365
39 Kappen M, van der Donckt J, Vanhollebeke G, Allaert J, Degraeve V, Madhu N, et al. Acoustic speech features in social comparison: how stress impacts the way you sound. Sci Rep. 2022;12(1):22022. https://doi.org/10.1038/s41598-022-26375-9
40 Rothkrantz LJM, Wiggers P, van Wees J-WA, van Vark RJ. Voice stress analysis. Text, speech and dialogue. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004; p. 449–56.
41 Mendoza E, Carballo G. Acoustic analysis of induced vocal stress by means of cognitive workload tasks. J Voice. 1998;12(3):263–73. https://doi.org/10.1016/s0892-1997(98)80017-9
42 Perrine BL, Scherer RC. Aerodynamic and acoustic voice measures before and after an acute public speaking stressor. J Speech Lang Hear Res. 2020;63(10):3311–25. https://doi.org/10.1044/2020_JSLHR-19-00252
43 Pisanski K, Nowak J, Sorokowski P. Individual differences in cortisol stress response predict increases in voice pitch during exam stress. Physiol Behav. 2016;163:234–8. https://doi.org/10.1016/j.physbeh.2016.05.018
44 Vásquez-Correa JC, Klumpp P, Orozco-Arroyave JR, Nöth E. Phonet: a tool based on gated recurrent neural networks to extract phonological posteriors from speech. Interspeech 2019. ISCA. ISCA; 2019. https://doi.org/10.21437/interspeech.2019-1405
45 Buchanan TW, Laures-Gore JS, Duff MC. Acute stress reduces speech fluency. Biol Psychol. 2014;97:60–6. https://doi.org/10.1016/j.biopsycho.2014.02.005
46 Belalcázar-Bolaños EA, Orozco-Arroyave JR, Vargas-Bonilla JF, Haderlein T, Nöth E. Glottal flow patterns analyses for Parkinson's disease detection: acoustic and nonlinear approaches. Text, speech, and dialogue. Cham: Springer International Publishing; 2016; p. 400–7.
47 Saslow LR, McCoy S, van der Löwe I, Cosley B, Vartan A, Oveis C, et al. Speaking under pressure: low linguistic complexity is linked to high physiological and emotional stress reactivity. Psychophysiology. 2014;51(3):257–66. https://doi.org/10.1111/psyp.12171