Discover Psychology

https://doi.org/10.1007/s44202-026-00679-7

Article in Press

Singing reveals vocal biomarkers of stress and cognitive load

Received: 17 November 2025

Accepted: 17 March 2026

Published online: 10 April 2026

Cite this article as: Rodrigo-Herrero I., Navas-León S. & Duñabeitia J.A. Singing reveals vocal biomarkers of stress and cognitive load. *Discov Psychol* (2026). https://doi.org/10.1007/s44202-026-00679-7

Irene Rodrigo-Herrero, Sergio Navas-León & Jon Andoni Duñabeitia

Manuscript Category: Brief Report

Title: Singing Reveals Vocal Biomarkers of Stress and Cognitive Load

Running title: Vocal biomarkers and singing

Authors: Irene Rodrigo-Herrero[1], Sergio Navas-León[1], Jon Andoni Duñabeitia[1*]

Affiliations:

[1] Centro de Investigación Nebrija en Cognición (CINC), Facultad de Lenguas y Educación, Universidad Nebrija, Madrid, Spain

***Correspondence:** Jon Andoni Duñabeitia. Email: jdunabeitia@nebrija.es

ORCIDs:

Irene Rodrigo-Herrero: https://orcid.org//0009-0008-5948-5571

Sergio Navas-León: https://orcid.org/0000-0002-0889-8252

Jon Andoni Duñabeitia: https://orcid.org/0000-0002-3312-8559

Declarations

Funding

This study has been partially supported by by grant "Accesibilidad lingüística en la web: investigación psicolingüística para la implementación de la Lectura Fácil" funded by Fundación Ramón Areces and by PID2024-161331NB-I00 funded by MICIU/AEI/10.13039/501100011033.

Conflict of interests/Competing interests

At the time of submission, Jon Andoni Duñabeitia was Guest Editor of the Collection to which this manuscript was submitted. Neither he nor any of the co-authors listed in this article had any involvement in the handling, peer review, or decision-making process for this manuscript.

Ethical approval

This study was approved by the Ethics Committee of Universidad Nebrija (Code: UNNE-2025-048). Data collection in this study followed the ethical standards laid down in the Declaration of Helsinki of 1964 and its subsequent amendments.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Data and Code availability

The materials, data and code are freely available at Open Science Framework (OSF). The materials can be accessed via https://osf.io/eg6xz/files/osfstorage/69a968ac5b49191c5459d1cb. The data can be accessed via https://osf.io/eg6xz/files/osfstorage/69a9680b0cb53e92ef79755f. The code can be accessed via https://osf.io/eg6xz/files/osfstorage/69a96860dea11e450f79753c.

ABSTRACT

Early detection of neurocognitive and mental health alterations is limited by the high costs and invasiveness of the protocols. Vocal biomarkers offer a non-invasive alternative, capturing prosodic and spectral features linked to cognitive and emotional states. Singing non-lexical syllabic vocalizations as a source for vocal biomarkers offers advantages over traditional speech-based methods, including language independence, scalability, and privacy preservation. This study evaluates whether singing melodies based on non-lexical syllables could provide reliable biomarkers sensitive to stress and cognitive load. Fourteen native Spanish-speaking participants completed a singing task under different stress conditions (acute stress vs. neutral) and different cognitive load conditions (immediate vs. delayed reproduction). Prosodic and spectral features were extracted, and results were explored in a series of linear mixed-effects models. Acute stress increased F0, decreased F1 and F2, and redistributed spectral energy (higher centroid, spread, flatness), producing a noisier output. Increased cognitive load led to shorter singing durations, increased F0, jitter and shimmer, and a flatter spectrum (lower kurtosis, higher flatness). Results demonstrated that stress engages autonomic arousal and articulatory changes, whereas cognitive load affects mainly control and stability. Shared markers index general arousal, while stress- and delay-specific features provide complementary sensitivity. Overall, singing non-lexical syllables stands as a feasible, low-burden, language-independent task for scalable vocal biomarker research. Alongside research on deep learning models, this opens new avenues for creating computer-based systems, either semi or fully automated, that can efficiently detect cognitive and emotional states.

Keywords: Vocal biomarkers; Acute stress; Cognitive load

Introduction

Despite substantial advances, the early detection of neurocognitive diseases and mental health conditions remains constrained by high costs, invasive procedures and reliance on specialized settings [1,2]. Biomarkers, conceived as objectively measurable indicators of biological processes, offer a promising avenue to stratify risk earlier and support population-based screening. They are being explored as tools for identifying individuals at elevated risk of neurocognitive diseases and for informing preventive interventions [1,3].

In this context, vocal biomarkers have gained attention as a cost-effective and non-invasive tool for detecting subtle neurocognitive and emotional changes that traditional assessments may miss [4,5]. They are derived from speech, capturing both acoustic and linguistic features linked to physiological and psychological states. Acoustic features include prosodic parameters such as pitch (i.e., fundamental frequency; hereafter, F0), formants, energy, jitter and shimmer, as well as spectral descriptors like flux, slope, centroid and entropy [6-8]. These acoustic features are increasingly recognized for their potential to reveal subtle neurocognitive and emotional changes. For instance, Kappen et al. [9] successfully differentiated between stressed and non-stressed participants by analyzing these features, and Ding et al. [10] identified individuals with cognitive impairment using similar acoustic-based approaches. Linguistic features, in turn, reflect aspects such as vocabulary richness and syntactic complexity [11,12]. However, linguistic features tend to be less transferable across tasks and populations, given their dependence on language structure and semantics [13]. By contrast, acoustic features are largely independent of language, show greater robustness and cross-linguistic generalizability, and are also sensitive to changes in cognitive load and emotional variation [14].

As said, the most relevant limitation of vocal biomarkers that prevents their generalized use in clinical settings is the lack of standardization due to language dependence [15,16]. In this context, singing melodies consisting of articulated syllabic vocalizations emerges as an unexplored but promising alternative. Since it is not dependent on language production, it can be adapted easily to different linguistic contexts, standardized through fixed contours, durations or prompts, and seamlessly integrated into standard devices for real-world applications [17,18]. Additionally, singing has been shown to be sensitive to changes in basic psychological processes such as attention or memory [19-21], suggesting that it could serve as an experimental paradigm acting as a source of vocal biomarkers that reflect changes in an individual's cognitive state [5].

The present study examines whether a brief singing task provides reliable prosodic and spectral biomarkers of cognitive alterations. To the best of our knowledge, this is the first study to evaluate singing of simple melodies as a vocal task with the potential to provide biomarkers that are scalable to large populations and generalize across contexts, which is an urgent call in the field as described above. To this end, two manipulations were incorporated. First, we explored the relevance of singing for detecting altered cognitive states by measuring induced acute stress. This approach is supported by increasing evidence that stress produces measurable changes in vocal features [22-24]. We predicted that the analysis of the acoustic biomarkers associated with an acute stress situation could be detected using recordings from a simple singing task completed immediately after a stress-inducing task. Second, we explored the sensitivity of this singing task to increased cognitive load by manipulating the moment at which participants had to repeat the melody that had to be hummed (i.e., either simultaneously, paired with the auditorily delivered template, or with a temporal delay, right after the template), a manipulation commonly used in speech shadowing to increase cognitive load by taxing working memory and attentional resources during vocal production. Seminal research showed that simultaneous or delayed repetition constitutes an on-line auditory tracking task that places high demands on attention and working memory, thereby increasing overall cognitive load [25-27]. Recent evidence from verbal shadowing and dual-task studies further indicates that increased cognitive load can interfere with the allocation of resources required for monitoring and controlling vocal production, even in the absence of stress-related autonomic activation [28-30]. In this regard, retaining a melody in memory over time has been shown to impose greater cognitive demands than immediate repetitions [31] and this delayed recall produce alterations in speech that correlate with early signs of cognitive decline [32,33]. In sum, we extracted a pre-specified set of acoustic features (prosodic and spectral) commonly used in speech-based biomarker research, and we assessed their sensitivity, stability and correspondence with speech-derived measures.

Methods

Participants

Participants were native Spanish speakers with normal hearing recruited from Universidad Nebrija. We excluded those with prior musical or singing training, those affected by unexpected technical issues during recording, and those undergoing treatment for stress or anxiety. In addition, participants completed the Positive and Negative Affect Schedule (PANAS-SF [39]) to evaluate their initial emotional state. Those with scores for negative affect higher than 29 and positive affect lower than 20 were considered to have an altered emotional state and were therefore excluded from the study. The initial sample consisted of 17 participants aged between 18 and 35 years. Three participants were removed due to unexpected technical problems during the recording process, resulting in a final sample of 14 participants with ages between 18 and 35 (M = 24.93, SD = 3.97; 11 females). Based on preceding evidence [34,35], we considered a conservative moderate-to-large effect size (β = 0.50) as a realistic estimate for our design. Simulation-based power analyses were conducted in R using lme4 for linear mixed-effects modelling [36] and simr for power estimation [37]. Power simulations (100 Monte Carlo runs) indicated that with 14 participants and 12 trials each (168 total observations), the current design achieved approximately 84-88% power to detect the expected effect. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments after obtaining approval from the institutional review board of Universidad Nebrija (Code: UNNE-2025-048). At the beginning of the experiment, written informed consent was obtained from all participants, who received a small monetary compensation. After participating in the experiment, people were debriefed.

Materials

For this study, two melodies (A and B) were created for participants. Both melodies had the same duration (19 seconds) and were matched as closely as possible on relevant acoustic parameters (e.g., pitch range, interval structure, rhythmic complexity). Each melody consisted of the continuous repetition of the same non-lexical syllable "la" (i.e., /la/), produced as a sequence of repeated syllables (e.g., "la-la-la") throughout the entire melodic contour. The fundamental frequency (F0) range was selected to be broadly comfortable for both male and female participants, and the same melodic templates were used for all participants without pitch-shifting by sex. Corresponding values are provided in **Table 1**. The melodies were synthesized using Praat [38], version 6.4.44) and standardized in loudness.

Table 1
Values of melodies A and B for each variable of interest.

Variable	Melody A	Melody B
Duration (sec)	19	19
Pitch (F0) (Hz)	260.89	269.60
Formant F1 (Hz)	671.74	660.25
Formant F2 (Hz)	1394.86	1470.57
Jitter (%)	5.4	5.8
Shimmer (%)	0.51	0.48
Tempo (BPM)	95.78	102.88

Energy (dBFS)	-24.20	-24.37
Spectral Centroid (Hz)	785.88	705.43
Spectral Spread (Hz)	630.30	674.31
Spectral Kurtosis	4.93	4.97
Spectral Skewness	2.98	3.03
Spectral Flatness	0.31	0.37
Spectral Flux	0.005	0.004

Procedure

The experiment was performed in a soundproof room. The tasks were performed using a 14-inch LED monitor with a refresh rate of 60 Hz and a resolution of 1920 x 1200 pixels. An AKG P420 microphone was used for recording and was connected to an M-Track Duo system to transfer the recording to a computer.

Participants first completed the Positive and Negative Affect Schedule (PANAS-SF; [39]). Next, each participant was assigned to one of four counterbalanced sequences, depending on the order of the stress-inducing conditions (stress task or neutral task) and the melody that was used (melody A or melody B, differentiated to avoid any memory effect between conditions) (see **Figure 1**). The experiment used a within-subject design, so that all participants completed both the stress and neutral task and listened to melody A and B, although in a different counterbalanced order. In both stress-inducing and neutral tasks, participants answered arithmetic and general knowledge questions. The stress task followed the original protocol of Almazrouei et al, (2023) [40], employing the Spanish adaptation developed by Navas-León and Duñabeitia [41] (see **Figure 2**). In this task, participants were told that their performance was being monitored and received immediate on-screen negative feedback, while this was not the case in the neutral version. In the neutral task, participants were shown questions of a very easy nature, while in the stress-inducing task, they were given questions of a more challenging nature.

After completing either, the stress-inducing or the neutral version of the task, participants were prompted with one of the two melodies (A or B; counterbalanced). During the subsequent task, they were presented with the alternate melody, ensuring that each melody was heard in a different task. After each of the main tasks (stress and neutral), participants completed six experimental trials: three trials in which they had to sing the melody immediately during the playback of the melody and three trials in which they had to wait until the melody had finished playing to start singing it. The presentation of the trials was randomized across participants. Participants were asked to vocally imitate the melody using the same repeated syllables provided in the models (i.e., “la-la-la”)**.** To avoid any spillover stress effects across tasks, immediately after completion of the stress task and the associated melody-repetition trials, participants completed an 11-minute guided meditation in which they had the option to relax while listening to a professional mindfulness podcast (CogniFit Inc., San Francisco, US). After each task, participants rated their perceived stress on a 0–100 visual analogue scale (VAS) (with a randomized slider start) and their task emotional valence on a 1–7 Likert scale [40-42].

The full procedure took approximately 50 minutes. The materials, scripts and data associated with this study are openly available at https://osf.io/eg6xz/overview?view_only=e66a4d15a2f94a4da421c595eea85993.

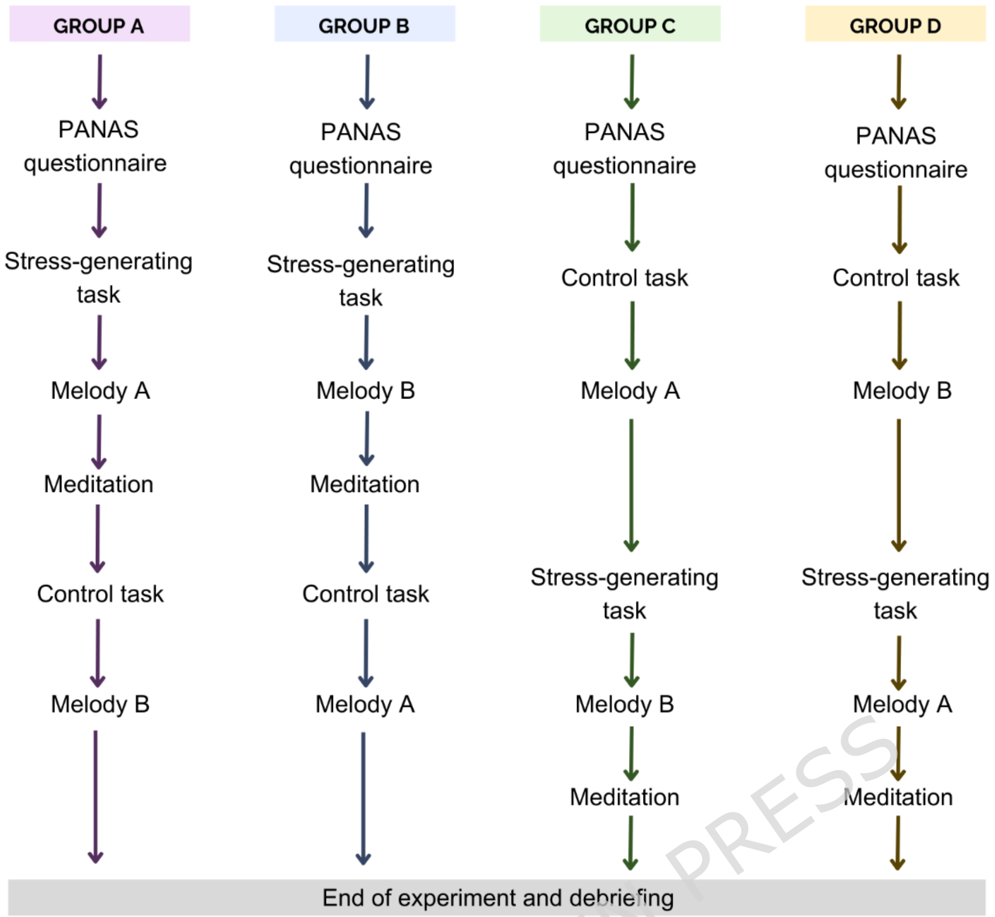


Figure 1. Schematic depiction of the experimental procedure.

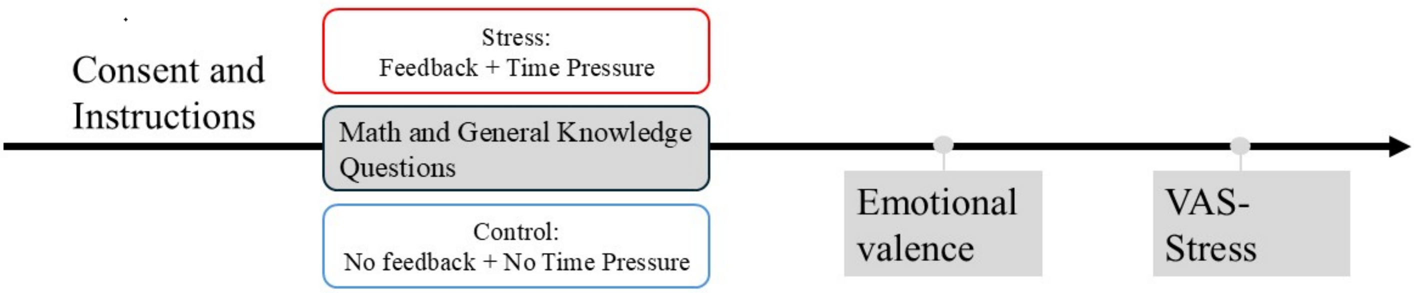


Figure 2. Graphic timeline of the stress-generating/control task (adapted from Almazrouei et al., 2023).

Recording and pre-processing

All auditory stimuli were delivered through closed-back over-ear Sennheiser HD 350BT headphones. The audio files were recorded using Praat, with a sampling frequency of 44100 Hz and mono channels. All the recordings were individually inspected to ensure the absence of clipping or excessive background noise and the resulting recordings were saved in WAV format. All audio files were pre-processed prior to feature extraction to ensure consistency and quality of the signals. Recordings were normalized in amplitude to avoid variability due to recording intensity [43]. For formant analysis, a pre-emphasis filter was applied to enhance high-frequency components, followed by linear predictive coding (LPC). To improve robustness, only valid LPC solutions were retained, and candidate formants were filtered to those above 90 Hz with bandwidths below 400 Hz [44]. In tempo estimation, spurious peaks were controlled by setting a minimum peak height and minimum inter-peak distance, preventing false detections.

The acoustic variables were selected based on prior evidence showing that stress and emotional arousal affect multiple dimensions of vocal production. Duration and tempo were included to capture stress-related temporal alterations in speech associated with increased cognitive load [45-47]. Energy was used to reflect changes in vocal intensity under heightened arousal [46,47] and mean fundamental frequency (F0) was included as a robust marker of stress-related pitch modulation [48-50]. Formant frequencies F1 and F2 serve as crucial indicators of vocal tract resonance and articulation pattern [51] and jitter and shimmer were included to evaluate the effects on voice control [52]. Spectral centroid, spectral spread, spectral skewness, and spectral kurtosis were used to characterize stress-related changes in vocal timbre and spectral energy distribution [46], while spectral flatness and spectral flux indexed increased noisiness and instability in the acoustic signal under stress [45,48].

Results

Stress levels and emotional valence

To assess task differences in stress-related responses, independent samples t-tests were conducted for stress and emotional valence.

Stress levels were significantly higher in the stress-inducing task ($M = 74.00$, $SD = 25.61$) than in the neutral task ($M = 18.93$, $SD = 17.67$). Since the assumptions of normality (Shapiro–Wilk $p = .48$) and homogeneity of variances were not violated, Student's t-test was used, revealing a statistically significant difference, $t(13) = -8.17$, $p < .001$, Cohen's $d = -2.18$, $SE = 0.55$. Similarly, emotional valence ratings were significantly lower in the stress-inducing task ($M = 2.93$, $SD = 1.39$) than in the neutral task ($M = 5.36$, $SD = 1.28$). The assumptions of normality (Shapiro–Wilk $p = .02$) and homogeneity of variances were violated. A Wilcoxon's signed-rank test confirmed a significant difference, $z = -3.06$, $p = .002$, $r = 1$, $SE = 0.32$.

These findings indicate that the stress-inducing task generated significantly greater stress levels and lower emotional valence than the neutral task, with large and consistent effect sizes across both measures.

Vocal biomarkers

The vocal parameters were extracted using Matlab version 25.1.0.2943329 (R2025a). Signal analysis and feature computation were carried out using specialized toolboxes (Signal Processing Toolbox, DSP System Toolbox, Audio Toolbox). All statistical analyses were conducted in R (version 4.2.3) via RStudio (R Core Team, 2018). Final values were computed as the difference between each participant's measured acoustic parameters and the corresponding reference values for each melody, and these difference scores were then standardized (z-scores) to facilitate comparisons across acoustic measures. Linear mixed-effects models (LMMs) were fitted using the lme4 [39], lmer Test [53] and afex [54] packages. LMMs were chosen to account for the nested structure of the data, which included repeated measurements within participants and across melodies. A top-down model selection strategy was followed [55,56]. For each dependent variable, a series of different models were fitted that systematically varied in their structure of fixed and random effects. While the initial models included the standardized result of the PANAS-SF (Positive Affect: M = 32.86, SD = 7.00; Negative Affect: M = 15.93, SD = 4.48) questionnaire as a covariate to test whether it moderated the effects of stress, this variable did not significantly improve any model's fit and was therefore excluded from the final models. The order of the trials was also analyzed as a covariate, but it did not significantly modify any of the effects, and it was therefore excluded since no significant learning effect was observed. Two possible fixed-effect structures were considered in the models: one that included the interaction between Timing (immediate vs. delayed) and State (neutral vs. stress) and one that did not. Each fixed-effect model was then paired with random-effect structures for Participant: random slopes for Timing, for State, or including no random slopes. All models included a random intercept for melody. Only models that significantly improved upon the null model were retained for final comparison. In cases where multiple significant models provided similar explanatory power, preference was given to the most parsimonious model (i.e., with fewer parameters and lower AIC). Marginal and conditional R^2 values were computed, along with intraclass correlation coefficients (ICC), to assess model fit and the proportion of variance attributable to random effects. For each dependent variable, the final selected model is described. Descriptive statistics for all conditions are reported in **Table 2.**

Table 2

Means and standard deviations of each variable for all four conditions

Variable	Neutral	Stress	Immediate	Delayed
Duration (sec)	17.91 ± 1.34	17.58 ± 1.54	18.5 ± 0.74	17 ± 1.6
Pitch (F0) (Hz)	215.84 ± 41.24	230.3 ± 41.36	218.91 ± 41.99	227.27 ± 41.46
Formant F1 (Hz)	792.66 ± 295.71	697.81 ± 124.93	742.81 ± 221.32	747.06 ± 241.09

Formant F2 (Hz)	1849.51 ± 313.6	1639.18 ± 334.61	1715.49 ± 301.37	1774.11 ± 373.84
Jitter (%)	1.1 ± 0.2	1.06 ± 0.18	1.05 ± 0.18	1.11 ± 0.19
Shimmer (%)	5 ± 1.81	5.12 ± 2.26	4.72 ± 1.96	5.39 ± 2.08
Tempo (BPM)	93.2 ± 5.53	94.33 ± 4.29	93.18 ± 4.97	94.36 ± 4.91
Energy (dBFS)	-61.69 ± 6.64	-60.25 ± 5.89	-60.44 ± 6.07	-61.49 ± 6.51
Spectral Centroid (Hz)	695.47 ± 172.21	732.43 ± 170.25	701.67 ± 188.5	734.46 ± 163.34
Spectral Spread (Hz)	1227.84 ± 305.43	1476.46 ± 522.43	1321.23 ± 466.63	1384.18 ± 423.17
Spectral Kurtosis	8.28 ± 4.66	7.78 ± 5.66	8.55 ± 5.72	7.51 ± 4.56
Spectral Skewness	2.2 ± 0.73	1.99 ± 0.76	2.2 ± 0.81	2 ± 0.68
Spectral Flatness	0.32 ± 0.08	0.35 ± 0.1	0.32 ± 0.09	0.35 ± 0.09
Spectral Flux	0.02 ± 0.02	0.02 ± 0.06	0.02 ± 0.06	0.02 ± 0.03

Duration

Duration was measured as the amount of time (in seconds) that participants spent repeating the melody. The best-fitting model included both fixed main effects (Timing and State) without their interaction and a random structure with random slopes for State by participant:

Value ~ Timing + State + (1 + State | Participant) + (1 | Melody)

A significant main effect of Timing emerged ($t = 8.76$, $p < .001$), whereas the effect of State was not significant ($t = 1.66$, $p = .12$). Participants produced shorter melody durations in the delayed compared to the immediate condition ($M_{immediate} = 18.5$ s vs. $M_{delayed} = 17.0$ s).

Pitch (F0)

Pitch reflects the F0 of vocal fold vibration. The best-fitting model included both fixed effects without their interaction and a random structure with random slopes for Timing by participant:

Value ~ Timing + State + (1 + Timing | Participant) + (1 | Melody)

Both effects of Timing and State were significant (t = 3,61, p = .0028, and t = 7.99, p > .001, respectively). Participants in the stress condition produced a higher pitch compared to the neutral condition ($M_{neutral}$ = 215.84 Hz vs. M_{stress} = 230.30 Hz). Similarly, F0 values were higher in the delayed than in the immediate condition ($M_{immediate}$ = 218.91 Hz vs. $M_{delayed}$ = 227.27 Hz).

Formant F1

Formant F1 corresponds to the first spectral peak resulting from vocal tract resonance. The best-fitting model included both fixed main effects without their interaction and a random structure with random slopes for Timing by participant:

Value ~ Timing + State + (1 + Timing | Participant) + (1 | Melody)

A significant main effect of State was observed (t = 3.48, p < .001), while Timing showed no significant effect (t = 0.8, p = .93). Specifically, F1 decreased in the stress condition ($M_{neutral}$ = 792.66 Hz vs. M_{stress} = 697.81 Hz).

Formant F2

Formant F2 represents the second spectral peak produced by vocal tract resonance. The best-fitting model included both fixed main effects without interaction and a random structure with no slopes:

Value ~ Timing + State + (1 | Participant) + (1 | Melody)

Participants in the stress condition showed significantly lower F2 values than those in the neutral condition (t = 4.36, p < .001; $M_{neutral}$ = 1849.51 Hz vs. M_{stress} = 1639.18

Hz). No significant differences were found between delayed and immediate conditions ($t = 1.39$, $p = .17$).

Jitter

Jitter measures frequency variability across cycles. The best-fitting model included both fixed effects with their interaction and a random structure with random slopes for Timing by participant:

*Value ~ Timing * State + (1 + Timing| Participant) + (1 | Melody)*

Although the interaction between State and Timing was not significant ($p = .1$), it was retained because it improved overall model fit. The effects of Timing ($t = 3.2$, $p = .005$), and and State ($t = 2.26$, $p = .002$) were significant. Jitter increased in the delayed compared to the immediate condition ($M_{immediate} = 1.05$ vs. $M_{delayed} = 1.11$). The non-significant interaction suggested that the effect of Timing (delayed vs. immediate) was attenuated under stress conditions.

Shimmer

Shimmer captures amplitude variability. The best-fitting model included both fixed main effects without interaction and a random structure with no slopes:

Value ~ Timing + State + (1 | Participant) + (1 | Melody)

The fixed effect of Timing was significant ($t = 4.72$, $p < .001$), whereas the effect of State was not ($t = 0.99$, $p = .32$). Shimmer increased significantly in the delayed compared to the immediate condition ($M_{immediate} = 4.72$ vs. $M_{delayed} = 5.39$).

Energy

Energy reflects the vocal signal’s intensity in dBFS (Decibel Full Scale), where 0 dBFS represents the maximum level. The best-fitting model included both fixed main effects without interaction and a random structure with no slopes:

Value ~ Timing + State + (1 | Participant) + (1 | Melody)

The fixed effects of State and Timing were significant ($t = 2.98$, $p < .01$, and $t = 2.14$, $p = .03$, respectively). Participants in the stress condition exhibited higher energy levels ($M_{neutral} = -61.69$ vs. $M_{stress} = -60.25$), and energy was lower in the delayed compared to the immediate condition ($M_{immediate} = -60.44$ vs. $M_{delayed} = -61.49$).

Tempo

Tempo represents the perceived speed or pacing of the performance. None of the tested models differed significantly from the null model, indicating no significant effects of either Timing or State, nor their interaction.

Spectral Centroid

The spectral centroid reflects the central frequency of the spectral energy distribution. The best-fitting model included both fixed effects with their interaction and a random structure with random slopes for Timing:

*Value ~ Timing ** State *+ (1 + Timing | Participant) + (1 | Melody)*

Although the interaction did not reach statistical significance ($p = .1$), it was retained because it improved model fit. The fixed effects of Timing and State were significant ($t = 2.7$, $p = .007$, and $t = 5.5$, $p < .001$, respectively). Spectral centroid was higher in the delayed compared to the immediate condition ($M_{immediate} = 701.67$ vs. $M_{delayed} = 734.46$), and it increased under stress ($M_{neutral} = 695.47$ vs. $M_{stress} = 732.43$). Timing×State interaction approached significance ($p = .060$). Simple effects analyses revealed that under neutral conditions, the spectral centroid was significantly lower in the delayed than in the immediate condition ($t = 2.73$, $p = .007$), whereas under stress this difference was negligible ($t = -0.08$, $p = .934$).

Spectral Spread

Spectral spread quantifies the dispersion of spectral energy around the centroid. The best-fitting model included both fixed main effects without interaction and a random structure with no slopes:

Value ~ Timing + State + (1 | Participant) + (1 | Melody)

A significant main effect of State emerged ($t = 5.06$, $p < .001$), whereas the effect of Timing was not significant ($t = 1.36$, $p = .17$). Spectral spread increased in the stress condition ($M_{neutral} = 1227.84$ vs. $M_{stress} = 1476.46$).

Spectral Skewness

Spectral skewness measures asymmetry in the spectral energy distribution. None of the tested models differed from the null model, indicating no significant effects for either Timing or State, nor interaction.

Spectral Kurtosis

Spectral kurtosis represents the peakedness or flatness of the spectral energy distribution. The best-fitting model included both fixed main effects without interaction and a random structure with no slopes:

Value ~ Timing + State + (1 | Participant) + (1 | Melody)

The fixed effect of Timing was significant ($t = 3.27$, $p = .001$), while the effect of State was not ($t = 1.46$, $p = .15$). Spectral kurtosis decreased in the delayed compared to the immediate condition ($M_{immediate} = 8.55$ vs. $M_{delayed} = 7.51$).

Spectral Flatness

Spectral flatness indicates the uniformity of power across frequencies. The best-fitting model included both fixed main effects without interaction and a random structure with no slopes:

Value ~ Timing + State + (1 | Participant) + (1 | Melody)

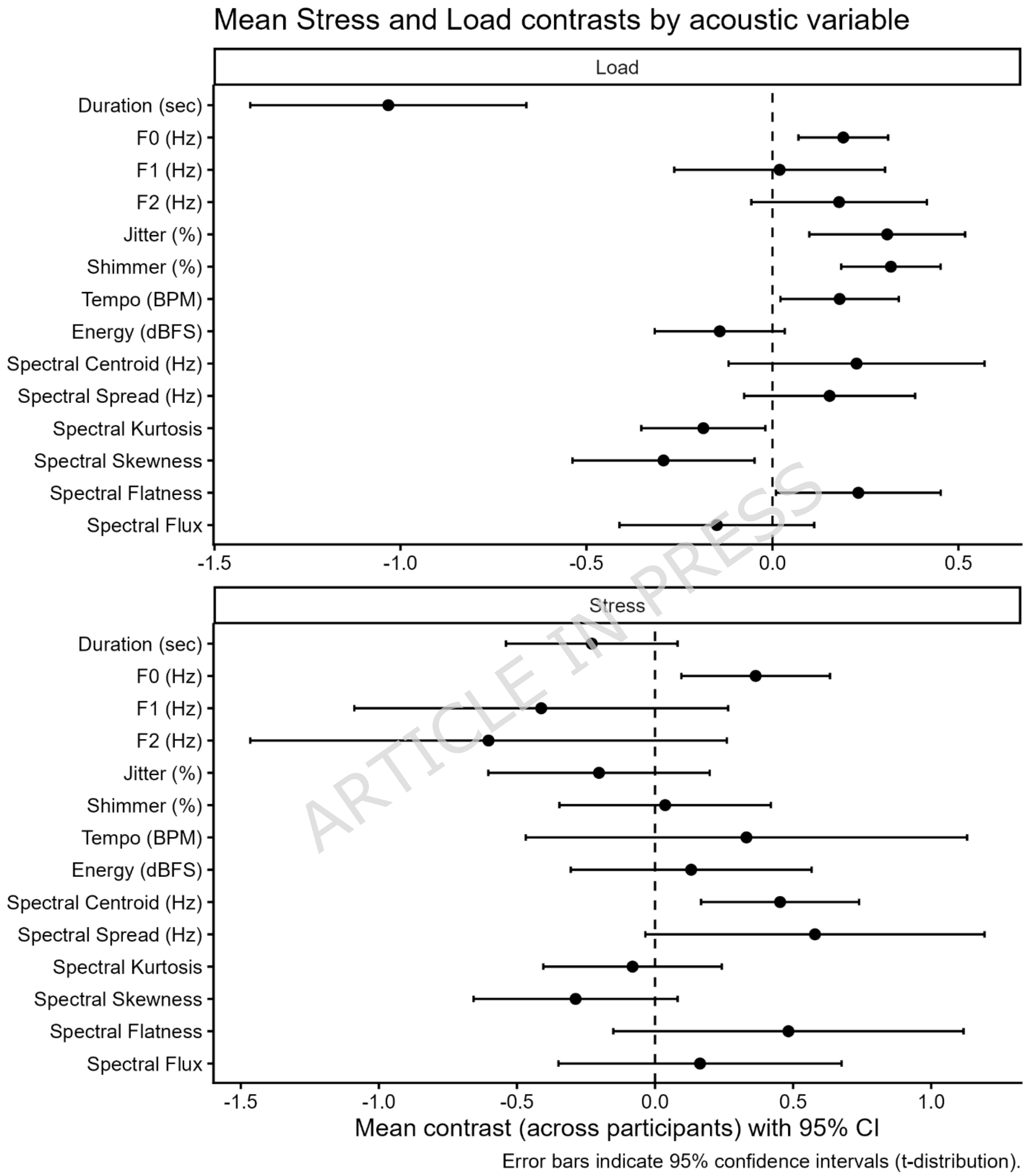
The fixed effects of State and Timing were significant ($t = 4.05$, $p < .001$, and $t = 2.33$, $p = .02$, respectively). Spectral flatness increased under stress ($M_{neutral} = 0.32$ vs. $M_{stress} = 0.35$) and in the delayed compared to the immediate condition ($M_{immediate} = 0.32$ vs. $M_{delayed} = 0.35$).

Spectral Flux

Spectral flux reflects the degree of rapid variation in spectral energy across time. None of the tested models differed from the null model, indicating no significant effects for either Timing or State, nor their interaction.

Figure 1

Estimated means of the difference between vocal acoustic parameters and reference melodies according to stress and cognitive load in z-score. Error bars indicate 95% confidence intervals



Discussion

The current study examined singing simple melodies as an effective source for vocal biomarkers. To this end, we evaluated the relation between stress and cognitive load, and several prosodic and spectral voice features extracted from singing. We hypothesized that singing could be sensitive to modulations of voice features aligned with those previously reported for speech. Singing samples were recorded under acute stress and cognitive load (induced by timing delay in melody repetition) manipulations. Overall, the results provided support our hypothesis, suggesting that sung non-lexical vocalizations are sensitive to stress and cognitive load and reflect them through measurable acoustic features.

An increase in stress activates the sympathetic nervous system (SNS), which produces muscle tension, increased heart rate and blood pressure, and bronchodilation [57], mediated via the vagus nerve pathway. These alterations affect vocal production, mainly F0 [50,58], fully consistent with the observed effects reported in the current study and in previous studies [9,56,57]. Results reflect the pattern observed in speech [9,59,60]. Accordingly, F1 and F2 decreased significantly under stress. These formants have been shown to be modulated by stress, either increasing or decreasing as a function of vowel identity: front vowels (i.e., /i/, /ɛ/) are associated with higher values, while back vowels (i.e., /o/ /a/) are associated with lower values [61]. Therefore, and considering that the melodies used in the current study were created by repeating the syllable /la/, our results are fully compatible with the decrease observed in earlier studies. Energy also increased under stress conditions, as also reported by Larsen et al. [22]. In the spectral domain, the spectral centroid and spread showed a significant increase under acute stress. Spectral flatness increased, from more tonal (neutral) to more noise-like (stress) sounds, indicating less uniform phonation [51]. Jitter decreased, while shimmer remained unchanged, consistent with preceding mixed evidence on the sensitivity of these features to stress [22,63,64]. Additionally, some studies have reported that the increase in muscle tension mediated by increased stress also creates less stable vocal fold patterns, resulting in increased jitter [52]. However, other studies suggest that compensatory vocal mechanisms lead to more regular vocal fold vibration despite increased laryngeal muscle tension, which could lead to a decreased jitter [51].

Finally, the duration of the vocalizations during the singing task was analyzed to assess potential memory effects of acute stress, given earlier evidence linking stress with reduced working memory [65]. However, no differences were observed. Nonetheless, stress effects on encoding have been shown to depend on age, with immediate recall being less impaired in younger adults [66], and the sample tested here was entirely made of undergraduate and graduate students.

Regarding the second manipulation of interest, cognitive load, the vocalizations of participants demonstrated sensitivity to the delayed condition in which they had to retain the melody in memory [26]. Participants produced shorter durations, consistent with lower selective attention and reduced memory selectivity under timing-related load [67]. F0, jitter and shimmer increased, suggesting reduced vocal stability due to a lack of control of vibrations of the vocal cords [49] (for similar effects, see Peters et al. [68]; Wynne, [69]). Formants F1 and F2 were not affected by the manipulations, which fits the view that cognitive load elicits a smaller autonomic response than the stress condition [70,71]. Our results are consistent with this explanation; F0 increased by 6.7% under stress conditions and 3.8% under the delayed condition, suggesting that, while effective, the cognitive load manipulation was subtler than the stress manipulation. In the spectral domain, kurtosis decreased, and flatness increased, indicating a more noise-like spectrum. Such a reduction in kurtosis under high cognitive load is consistent with previously reported flatter spectral distributions in similar contexts [72]. Spectral centroid showed a small significant increase, consistent with previous studies [73], only in neutral conditions, while spectral spread showed no reliable effect, warranting further investigation.

Several features remained unchanged under both stress and cognitive load. Tempo, for instance, showed no effect, mirroring contradictory speech findings [33,71-73].

Likewise, spectral skewness remained unchanged, and spectral flux, a measure of sudden changes in the frequency energy distribution, was also unaffected, likely reflecting participants trying to sustain stable notes as a strategy to control their vocal output. Taken together, the pattern points to two partly overlapping pathways. The stress condition resulted in higher autonomic arousal with articulatory reconfiguration and increased respiratory variability [70,71,77], eliciting a less tonal, more noise-like output. The delayed condition primarily taxed control and stability [52], with a decrease in total respiratory variability associated with top-down regulation [77].

Limitations and future lines of research

Some limitations should be acknowledged. First, the sample was relatively small (N = 14) and recruited by convenience, which reduces power for small effects and limits generalizability. Although the study was sufficiently powered to detect medium-to-large effects, the sample size may limit the precision of the LMM estimates and increase the risk of overfitting in multivariate modelling. Future research should therefore recruit larger and more diverse samples (ideally using stratified random sampling) to support more robust LMM estimation and to test multivariate models to better evaluate the biomarker potential of these acoustic measures. Second, the vocal task used only two melodies and a single syllabic vowel (/la/), restricting acoustic variability and articulatory breadth. Relatedly, the cognitive-load manipulation contrasted with concurrent versus delayed repetition; while this provides a clear contrast between externally guided and memory-based reproduction, concurrent playback necessarily involves simultaneous perception and production. Future work should therefore consider parametric delays (shorter vs. longer, without concurrent playback) to more cleanly isolate memory load. Lastly, the study relied exclusively on subjective and behavioral measures, without incorporating physiological indicators of stress (e.g., salivary cortisol, heart rate variability, or skin conductance), which limits the ability to characterize acute stress reactivity comprehensively. Incorporating physiological measures in future studies would strengthen convergent validity and help clarify discrepancies across stress-related markers.

Considering the current results, potential applications of singing as a source of vocal biomarkers emerge. A brief and standardized singing task could be deployed on commodity devices (smartphones, laptops) using fixed prompts (“sing now” vs. “sing after playback”) to monitor different cognitive states and conditions. The analysis of a compact vocal feature set obtained from such tasks could support a low burden monitoring tool in clinics, workplaces, or at home. These characteristics facilitate the application of deep learning models and other advanced machine learning techniques to analyze complex acoustic patterns, enabling the creation of semi-automated or fully automated systems capable of continuously monitoring cognitive function, supporting early detection, longitudinal monitoring, and personalized interventions. To this end, and to endorse clinical validation, future studies should orient efforts to gathering population-level data to establish normative reference values. All in all, the current study provides the first known piece of evidence demonstrating the value of a generalizable singing task as a source of vocal biomarkers that could be applied to different populations regardless of their language.

References

1. Hansson, O. (2021). Biomarkers for neurodegenerative diseases. *Nature Medicine*, 27(6), 954–963.
2. Santos, F., Cabreira, V., Rocha, S., & Massano, J. (2023). Blood Biomarkers for the Diagnosis of Neurodegenerative Dementia: A Systematic Review. *Journal of Geriatric* Psychiatry *and Neurology*, 36(4), 267–281.
3. Ritsner, M. S., & Gottesman, I. I. (2009). Where Do We Stand in the Quest for Neuropsychiatric Biomarkers and Endophenotypes and What Next? In M. S. Ritsner (Ed.), The Handbook of Neuropsychiatric Biomarkers, Endophenotypes and Genes*:*

Neuropsychological Endophenotypes and Biomarkers (pp. 3–21). Springer Netherlands.

4. Fagherazzi, G., Fischer, A., Ismael, M., & Despotovic, V. (2021). Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. *Digital Biomarkers*, 5(1), 78–88.
5. Mahon, E., & Lachman, M. E. (2022). Voice biomarkers as indicators of cognitive changes in middle and later adulthood. *Neurobiology of* Aging, 119, 22–35.
6. Ettore, E., Müller, P., Hinze, J., Riemenschneider, M., Benoit, M., Giordana, B., Postin, D., Hurlemann, R., Lecomte, A., Musiol, M., Lindsay, H., Robert, P., & König, A. (2023). Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review. JMIR *Mental* Health, 10, e37225.
7. Maffei, M. F., Green, J. R., Murton, O., Yunusova, Y., Rowe, H. P., Wehbe, F., Diana, K., Nicholson, K., Berry, J. D., & Connaghan, K. P. (2023). Acoustic Measures of Dysphonia in Amyotrophic Lateral Sclerosis. *Journal of* Speech*, Language, and Hearing Research*, 66(3), 872–887.
8. Shinohara, S., Toda, H., Nakamura, M., Omiya, Y., Higuchi, M., Takano, T., Saito, T., Tanichi, M., Boku, S., Mitsuyoshi, S., So, M., Yoshino, A., & Tokuno, S. (2020). Evaluation of the Severity of Major Depression Using a Voice Index for Emotional Arousal. Sensors, 20(18), 5041.
9. Kappen, M., van der Donckt, J., Vanhollebeke, G., Allaert, J., Degraeve, V., Madhu, N., Van Hoecke, S., & Vanderhasselt, M.-A. (2022). Acoustic speech features in social comparison: how stress impacts the way you sound. *Scientific Reports*, 12(1), 22022.
10. Ding, H., Lister, A., Karjadi, C., Au, R., Lin, H., Bischoff, B., & Hwang, P. H. (2024). Detection of Mild Cognitive Impairment From Non-Semantic, Acoustic Voice Features: The Framingham Heart Study. JMIR Aging, 7, e55126. https://doi.org/10.2196/55126

11. Gumus, M., Koo, M., Studzinski, C. M., Bhan, A., Robin, J., & Black, S. E. (2024). Linguistic changes in neurodegenerative diseases relate to clinical symptoms. *Frontiers in Neurology*, 15.

12. Robin, J., Harrison, J. E., Kaufman, L. D., Rudzicz, F., Simpson, W., & Yancheva, M. (2020). Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. *Digital Biomarkers*, 4(3), 99–108.

13. Huang, L., Yang, H., Che, Y., & Yang, J. (2024). Automatic speech analysis for detecting cognitive decline of older adults. *Frontiers in* Public Health, 12.

14. Menne, F., Dörr, F., Schräder, J., Tröger, J., Habel, U., König, A., & Wagels, L. (2024). The voice of depression: speech features as biomarkers for major depressive disorder. BMC Psychiatry, 24(1).

15. Ghasemzadeh, H., Hillman, R. E., & Mehta, D. D. (2024). Toward Generalizable Machine Learning Models in Speech, Language, and Hearing Sciences: Estimating Sample Size and Reducing Overfitting.

16. Parola, A., Rybner, A., Jessen, E. T., Damsgaard Mortensen, M., Larsen, S. N., Simonsen, A., Zhou, Y., Koelkebeck, K., Bliksted, V., & Fusaroli, R. (2023). Vocal markers of schizophrenia: assessing the generalizability of machine learning models and their clinical applicability.

17. Briganti, G., & Lechien, J. R. (2025). Voice Quality as Digital Biomarker in Bipolar Disorder: A Systematic Review. In Journal of Voice. Elsevier Inc.

18. Greenspon, E. B., & Montanaro, V. (2023). Singing ability is related to vocal emotion recognition: Evidence for shared sensorimotor processing across speech and music. *Attention, Perception, and Psychophysics*, 85(1), 234–243.

19. Halpern, A. R., & Pfordresher, P. Q. (2022). What do less accurate singers remember? Pitch-matching ability and long-term memory for music. *Attention, Perception, & Psychophysics*, 84(1), 260–269.

20. Weiss, M. W., & Peretz, I. (2019). Ability to process musical pitch is unrelated to the memory advantage for vocal music. Brain *and Cognition*, 129, 35–39.

21. Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the Way She Sings. *Psychological Science*, 23(10), 1074–1078.

22. Larsen, E., Murton, O., Song, X., Joachim, D., Watts, D., Kapczinski, F., Venesky, L., & Hurowitz, G. (2024). Validating the efficacy and value proposition of mental fitness vocal biomarkers in a psychiatric population: prospective cohort study. *Frontiers in* Psychiatry, 15.

23. Menne, F., Lindsay, H., Tröger, J., Paulmann, S., König, A., Steinbach, N., Reif, A., Plichta, M. M., & Schmidt-Kassow, M. (2025). Voice as objective biomarker of stress: association of speech features and cortisol. Acta *Neuropsychiatrica*, 37, e84.

24. Namkung, J., Kim, S. M., Cho, W. I., Yoo, S. Y., Min, B., Lee, S. Y., Lee, J.-H., Park, H., Baik, S., Yun, J.-Y., Kim, N. S., & Kim, J.-H. (2024). Novel Deep Learning-Based Vocal Biomarkers for Stress Detection in Koreans. Psychiatry *Investigation*, 21(11), 1228–1237.

25. Carey PW. Verbal retention after shadowing and after listening. *Percept Psychophys.* 1971;9(1):79–83. doi:10.3758/BF03213032.

26. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am.* 1953;25:975–979. doi:10.1121/1.1907229.

27. Lambert S. Shadowing. *Meta.* 1992;37(2):263–273. doi:10.7202/003378ar.

28. Hamada Y. Shadowing: Who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Lang Teach Res.* 2016;20(1):35–52. doi:10.1177/1362168815597504.

29. Meilinger T, Bülthoff HH. Verbal shadowing and visual interference in spatial memory. *PLoS One.* 2013;8(9):e74177. doi:10.1371/journal.pone.0074177.

30. Stevens CJ, Gibert G, Leung Y, Zhang Z. Evaluating a synthetic talking head using a dual task: Modality effects on speech understanding and cognitive load. *Int J Hum Comput Stud.* 2013;71(4):440–454. doi:10.1016/j.ijhcs.2012.12.003.

31. Sisakhti, M., Sachdev, P. S., & Batouli, S. A. H. (2021). The Effect of Cognitive Load on the Retrieval of Long-Term Memory: An fMRI Study. *Frontiers in Human Neuroscience*, 15.

32. Bae, M., Seo, M.-G., Ko, H., Ham, H., Kim, K. Y., & Lee, J.-Y. (2023). The efficacy of Memory load on speech-based detection of Alzheimer's disease. *Frontiers in* Aging *Neuroscience*, 15.

33. Young, C. B., Smith, V., Karjadi, C., Grogan, S.-M., Ang, T. F. A., Insel, P. S., Henderson, V. W., Sumner, M., Poston, K. L., Au, R., & Mormino, E. C. (2024). Speech patterns during memory recall relates to early tau burden across adulthood. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 20(4), 2552–2563.

34. Buchanan, T. W., Laures-Gore, J. S., & Duff, M. C. (2014). Acute stress reduces speech fluency. *Biological Psychology*, 97, 60–66.

35. Oesten, M., Richer, R., Abel, L., Rohleder, N., & Eskofier, B. M. (2023). VoStress – Voice-based Detection of Acute Psychosocial Stress. *2023* IEEE *EMBS International Conference on Biomedical and Health Informatics (*BHI*)*, 1–4.

36. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1).

37. Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods *in Ecology and Evolution*, 7(4), 493–498.

38. Boersma, P., & Weenink, D. (2025). Praat: Doing phonetics by computer (Version 6.4.44) [Computer software]. Retrieved September 25, 2025, from https://praat.org

39. Díaz-García, A., González-Robles, A., Mor, S., Mira, A., Quero, S., García-Palacios, A., Baños, R. M., & Botella, C. (2020). Positive and Negative Affect Schedule (PANAS): Psychometric properties of the online Spanish version in a clinical sample with emotional disorders. BMC Psychiatry, 20(1).

40. Almazrouei MA, Morgan RM, Dror IE. A method to induce stress in human subjects in online research environments. Behav Res Methods. 2023;55(5):2575-2582. doi:10.3758/s13428-022-01915-3.

41. Navas-León S, Duñabeitia JA. Validating an acute stress induction method in adult Spanish-speaking populations. Manuscript under review. 2026. doi:10.31234/osf.io/s8yvt_v2.

42. Koncz A, Köteles F, Aranyossy B, Horváth Á. Acute psychological stress does not influence joint position reproduction performance in the elbow joint. PLoS One. 2025;20(3):e0319061. doi:10.1371/journal.pone.0319061.

43. Alim, A., Purnapatra, S., Khondkar, M. J. A., Schuckers, S., & Imtiaz, M. H. (2024). Open-Source Pipeline for Noise-Resilient Voice Data Preparation.

44. Kanty, D., & Staroniewicz, P. (2025). Tests of the Influence of DAF (Delayed Auditory Feedback) on Changes in Speech Signal Parameters. *Applied Sciences*, 15(13), 7524.

45. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol Bull.* 2003;129(5):770–814. doi:10.1037/0033-2909.129.5.770.

46. Scherer KR. Vocal communication of emotion: A review of research paradigms. *Speech Commun.* 2003;40(1–2):227–256. doi:10.1016/S0167-6393(02)00084-5.

47. Schewski L, Doss MM, Beldi G, Keller S. Measuring negative emotions and stress through acoustic correlates in speech: A systematic review. *PLoS One.* 2025;20(7):e0328833.

48. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol.* 1996;70(3):614–636. doi:10.1037/0022-3514.70.3.614.

49. Laukka P, Linnman C, Åhs F, Pissiota A, Frans Ö, Faria V, et al. In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *J Nonverbal Behav.* 2008;32(4):195–214. doi:10.1007/s10919-008-0055-9.

50. Veiga DL, Almeida TM, Uchida RR, Cordeiro Q. The fundamental frequency of voice as a potential stress biomarker: A systematic review and meta-analysis. *Stress Health.* 2025;41(5):e70112.

51. Wang Q, Xu F, Wang X, Wu S, Ren L, Liu X. How anxiety state influences speech parameters: A network analysis study from a real stressed scenario. *Brain Sci.* 2025;15(3):262.

52. Teixeira J, Oliveira C, Lopes C. Vocal acoustic analysis – jitter, shimmer and HNR parameters. *Procedia Technol.* 2013;9:1112–1122. doi:10.1016/j.protcy.2013.12.124.

53. Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13).

54. Singmann, H., Kellen, D., Cox, G. E., Chandramouli, S., Davis-Stober, C. P., Dunn, J. C., Gronau, Q. F., Kalish, M., McMullin, S. D., Navarro, D., & Shiffrin, R. (2021). Statistics in the Service of Science*: Don't let the Tail Wag the Dog*.

55. Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.

56. Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092.

57. Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: a review. *Journal of voice : official journal of the Voice Foundation*, *27*(3), 390.e21–390.e3.9E29.

58. Zhang, Z. (2016). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, 140(4), 2614.

59. Godin, K. W., & Hansen, J. H. L. (2015). Physical task stress and speaker variability in voice quality. EURASIP *Journal on* Audio*,* Speech*, and* Music *Processing*, 2015(1), 29.

60. Tang, F., Chen, J., Dodge, H. H., & Zhou, J. (2022). The Joint Effects of Acoustic and linguistic markers for Early Identification of Mild Cognitive Impairment. *Frontiers in Digital Health*, *3*.

61. Qiu, C., & Liang, J. (2015). The Effect of stress on vowel space in Daxi Hakka Chinese. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2015-*January*, 973–977.

62. Alías, F., Socoró, J., & Sevillano, X. (2016). A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Applied Sciences*, 6(5), 143.

63. Pisanski, K., & Sorokowski, P. (2021). Human Stress detection: Cortisol Levels in Stressed Speakers Predict Voice-Based Judgments of Stress. Perception, 50(1), 80–87.

64. Tavi, L. (2017). Acoustic correlates of female speech under stress based on /i/-vowel measurements. *The International Journal of Speech, Language and the* Law, 24(2), 227–241.

65. Geißler, C. F., Friehs, M. A., Frings, C., & Domes, G. (2023). Time-dependent Effects of acute stress on working memory performance: A systematic review and hypothesis. Psychoneuroendocrinology, 148, 105998.

66. Hidalgo, V., Almela, M., Villada, C., & Salvador, A. (2014). Acute stress impairs recall after interference in older people, but not in young people. *Hormones and Behavior*, 65(3), 264–272.

67. Dubravac, M., & Meier, B. (2023). Cognitive load at encoding hurts memory selectivity. *Quarterly Journal of Experimental Psychology*, 76(7), 1515–1538.

68. Peters, J., Frank, M., & Rohloff, T. (2023). Effects of Cognitive Load on Vocal Fold Vibratory Patterns in Bilingual Speakers of Low and High German. *Journal of Voice*.

69. Wynne, H. (2011). The Effects of Cognitive Load on Language Production*: A Case Study of Aviation Communication*.

70. Ahmadi, N. K., Ozgur, S. F., & Kiziltan, E. (2024). Evaluating the Effects of Different Cognitive Tasks on Autonomic Nervous System Responses: Implementation of a High-Precision, Low-Cost Complementary Method. *Brain and Behavior*, *14*(10), e70089.

71. Won, E., & Kim, Y.-K. (2016). Stress, the Autonomic Nervous System, and the immune-kynurenine pathway in the Etiology of Depression. *Current Neuropharmacology*, 14, 665–673.

72. Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *The Journal of the Acoustical Society of America*, 93(5), 2962–2973.

73. Le, P. N., Ambikairajah, E., Epps, J., Sethu, V., & Choi, E. H. C. (2011). Investigation of Spectral centroid features for cognitive load classification. Speech *Communication*, 53(4), 540–551.

74. Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, 94, 166–176.

75. Çorlu, M., Maes, P.-J., Muller, C., Kochman, K., & Leman, M. (2015). The impact of Cognitive load on operatic singers’ timing performance. *Frontiers in Psychology*, 6.

76. Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related Cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications*, *3*(1), 46.

77. Vlemincx, E., Taelman, J., De Peuter, S., Van Diest, I., & Van den Bergh, O. (2011). Sigh rate and respiratory variability during mental load and sustained attention. Psychophysiology, 48(1), 117–120.