The title

First Author[1] & Ernst-August Doelle[1,2]

[1] Wilhelm-Wundt-University

[2] Konstanz Business School

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. First Author: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Ernst-August Doelle: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to First Author, Postal address. E-mail: my@email.com

Abstract

Personality traits showed domain- and parameter-specific moderation: Negative Affectivity amplified stress-induced F0 increases, while Psychoticism uniquely modulated voice quality (NNE) during recovery

Keywords: keywords

Word count: X

The title

Introduction

Personality pathology is widely theorized to manifest context-dependently, with maladaptive trait expression amplified, attenuated, or qualitatively altered by specific environmental demands (Hopwood et al., 2022; Wright & Simms, 2016). Yet empirical methods for capturing this dynamic interplay between person and situation in naturalistic settings remain underdeveloped. While it is broadly acknowledged that "context matters" for understanding personality pathology, the field has struggled to move beyond this truism toward rigorous operationalization and measurement of contextual processes as they unfold in daily life (Wright et al., 2019). Traditional personality assessment treats traits as decontextualized dispositions measured at single timepoints, ignoring the temporal dynamics and situational contingencies through which pathology is actually expressed. This gap between theory—which emphasizes person × environment transactions—and method—which relies on static, acontextual measurement—constrains both our understanding of personality pathology and our ability to predict when, where, and for whom maladaptive patterns will emerge.

Ecological momentary assessment (EMA) has partially addressed this limitation by capturing self-reported states and behaviors in real-world contexts, yielding valuable insights into within-person variability and situational reactivity (Trull & Ebner-Priemer, 2020). However, EMA remains fundamentally limited by its reliance on self-report. Individuals may lack introspective access to physiological stress responses, display biased recall or reporting, and experience assessment reactivity wherein the act of self-monitoring itself alters the phenomena under study (Barta et al., 2012). Moreover, repeated self-report imposes participant burden that constrains sampling density, limiting temporal resolution precisely when it is most needed—during acute stress episodes that unfold over minutes to hours. These constraints motivate the search for passive sensing approaches that can objectively, unobtrusively, and continuously capture psychological states and trait-by-situation interactions without requiring conscious self-evaluation.

Voice acoustics offer a promising solution. Vocal production is inherently psychophysiological: the acoustic properties of speech reflect underlying autonomic arousal, emotional state, and motor control processes that are partially outside conscious awareness (Scherer, 2003). Fundamental frequency (F0)—the acoustic correlate of perceived pitch—is modulated by laryngeal muscle tension, which increases systematically under stress and anxiety as sympathetic activation tightens vocal folds (Giddens et al., 2013). Voice quality parameters such as jitter (cycle-to-cycle pitch variability) and normalized noise energy (NNE, indexing glottal noise) capture phonatory control and vocal tract configuration, which are sensitive to both acute emotional arousal and chronic psychological states (Scherer et al., 2013). Formant frequencies, reflecting articulatory precision and vocal tract shape, provide additional markers of motor control stability that may be disrupted under stress or in individuals with executive dysfunction (Kent & Kim, 2003). Critically, these acoustic features can be extracted from brief, naturalistic speech samples—such as sustained vowel phonations or sentence reading—making voice a scalable, low-burden method for ambulatory assessment.

The theoretical relevance of vocal stress reactivity to personality pathology emerges from transactional models of person-situation interaction. Mischel and Shoda's (1995) Cognitive-Affective Personality System (CAPS) framework conceptualizes personality not as fixed traits but as conditional if-then behavioral signatures: stable individual differences in how people respond to specific situational features. Rather than asking whether someone is "high" or "low" on a trait in the abstract, CAPS emphasizes context-contingent patterns—if situation X, then response Y—wherein the strength and form of Y varies systematically across individuals as a function of personality. Similarly, Fleeson's (2001) whole trait theory distinguishes between density distributions (mean trait levels) and contingency patterns (situation-specific reactivity), arguing that comprehensive personality assessment must capture both. Applied to personality pathology, these frameworks predict that maladaptive traits should not merely correlate with outcomes but should moderate the impact of environmental stressors, shaping physiological and behavioral responses in domain-specific ways.

Vulnerability-stress models provide a complementary lens. If personality pathology dimensions reflect stable individual differences in stress sensitivity—heightened reactivity to threat, impaired recovery from challenge, or dysregulated arousal—then acute stressors should disproportionately affect individuals scoring high on relevant traits (Bolger & Zuckerman, 1995). The Personality Inventory for DSM-5 (PID-5; Krueger et al., 2012) operationalizes personality pathology as five dimensional traits that map onto theoretical constructs with clear stress-relevant mechanisms. Negative Affectivity captures anxiety sensitivity, emotional lability, and heightened threat reactivity—traits that should amplify physiological arousal under stress. Detachment reflects social withdrawal and anhedonia, potentially impairing post-stressor recovery through reduced social buffering and sustained rumination. Antagonism encompasses manipulativeness and callousness, which may paradoxically facilitate stress resilience by reducing empathic distress or may prolong tension through interpersonal conflict. Psychoticism indexes perceptual and cognitive dysregulation that could manifest in unstable vocal production even at baseline. Disinhibition reflects poor impulse control and affective instability, which may alter articulatory precision through compromised motor planning.

Despite this theoretical richness, empirical tests of how personality pathology traits moderate physiological stress responses remain scarce, particularly in naturalistic settings. Laboratory studies using standardized stressors offer experimental control but sacrifice ecological validity: artificially induced stress may not engage the same psychological processes as real-world evaluative threats. Conversely, field studies capturing daily stress typically rely on self-report measures vulnerable to the biases noted above. What is needed is an approach that combines (a) naturalistic stressors with genuine stakes for participants, (b) objective physiological markers not dependent on introspection, and (c) intensive longitudinal assessment to model within-person dynamics alongside between-person moderation.

The current study addresses this need through a multimodal ambulatory assessment design. We capitalized on the university examination period as an ecologically valid acute stressor with genuine evaluative significance. Participants (N = 141 female university students) completed the full 220-item PID-5 at baseline, then engaged in intensive EMA over 2.5 months using a brief 15-item version (three items per domain, factor-analytically selected to maximize domain representativeness). Vocal samples—sustained phonations of vowels /a/, /i/, and /u/—were recorded at three timepoints: baseline (distant from exams), pre-exam (the day before a major course examination), and post-exam (the day after completion). We extracted established acoustic features indexing arousal (fundamental frequency), voice quality (jitter, normalized noise energy), and articulatory precision (formant frequencies and variability). Bayesian hierarchical models with weakly informative priors tested whether PID-5 domains measured via EMA moderated stress-induced changes in vocal acoustics, distinguishing between acute stress reactivity (baseline $\rightarrow$ pre-exam) and post-stressor recovery (pre-exam $\rightarrow$ post-exam).

This design yields several innovations. First, it operationalizes context through a proximal, time-limited achievement stressor rather than retrospective self-reports of daily hassles, ensuring temporal precision in linking stressor exposure to vocal responses. Second, it employs passive acoustic sensing to capture physiological stress markers continuously and objectively, addressing self-report biases inherent in traditional EMA while maintaining ecological validity. Third, it integrates trait measurement through both comprehensive baseline assessment (220 items, single administration) and intensive repeated measurement (15 items, multiple occasions), enabling model comparison to evaluate whether brief EMA-based personality indicators provide equivalent predictive accuracy to traditional comprehensive questionnaires. Fourth, it tests theoretically grounded, domain-specific hypotheses about which personality pathology dimensions should moderate which aspects of vocal stress responses, moving beyond generic trait × stress interactions toward mechanistically informed predictions.

Our primary hypotheses concerned moderation of acute stress reactivity. We predicted that Negative Affectivity would amplify stress-induced increases in fundamental frequency, reflecting heightened autonomic arousal and anxiety sensitivity. We predicted that Detachment would impair post-stressor vocal recovery, reflecting sustained negative affect and reduced social buffering. We predicted that Antagonism might facilitate rapid recovery through reduced empathic distress or emotional contagion, though this prediction was exploratory given limited prior literature. We anticipated that Psychoticism and Disinhibition might show stable main effects on baseline vocal production—reflecting chronic dysregulation of phonatory and articulatory systems—rather than context-specific moderation of stress responses. Regarding voice quality and articulatory features, we remained agnostic: while theory suggests that stress should degrade phonatory control (increased jitter, elevated glottal noise) and articulatory precision (increased formant variability), it was unclear whether these effects would be uniformly expressed or moderated by specific personality pathology dimensions.

Methodologically, we anticipated that intensive EMA-based personality assessment would yield comparable predictive accuracy to comprehensive baseline assessment when predicting vocal stress responses. If brief indicators assessed repeatedly across multiple occasions capture core trait variance through explicit measurement error modeling—separating stable between-person differences from occasion-specific fluctuations—then EMA should approximate the predictive performance of extensive single-administration questionnaires despite using far fewer items. This hypothesis, if supported, would validate the efficiency of ambulatory trait assessment for research contexts where participant burden constrains comprehensive measurement.

By integrating passive acoustic sensing with intensive ambulatory trait assessment in a naturalistic stress context, this study advances methodological innovation in personality pathology research. It demonstrates that voice acoustics can serve as an objective, low-burden marker of both trait-level individual differences and context-dependent stress responses. It tests whether maladaptive personality traits operate as vulnerability factors that amplify physiological reactivity to environmental demands, providing empirical grounding for transactional models of personality pathology. And it contributes to the broader agenda of moving personality science beyond static, acontextual assessment toward capturing the dynamic, situated processes through which pathology manifests in daily life.

Methods

Participants

We recruited 141 female university students (M_age = [DA INSERIRE], SD = [DA INSERIRE]) from the University of Florence through course announcements and online postings. The sample was restricted to female participants to control for sex-related variation in vocal pitch characteristics, which could obscure personality-related effects and reduce statistical power. Fundamental frequency (F0) differs substantially between males and females (approximately 100 Hz lower in males) due to anatomical differences in vocal fold length and mass, making direct comparison problematic without large samples or complex statistical controls.

All participants were native Italian speakers with no reported history of voice disorders or current respiratory illness. Exclusion criteria included: (1) current or past psychiatric disorders requiring treatment, (2) substance use disorders, (3) self-reported hearing impairments that could affect voice monitoring, and (4) professional voice training (e.g., singing lessons), which could alter baseline vocal characteristics. Participants provided written informed consent and received course credit for participation. The study was approved by the University of Trieste Ethics Committee (protocol #[INSERIRE]).

Note on parallel investigation. A parallel study with male participants (N = 36) using an identical protocol examined MFCC-based patterns and stress-induced vocal alterations with focus on different acoustic features. The current manuscript reports exclusively on the female sample to maximize statistical power for examining personality moderation effects and ensure interpretability of pitch-related findings.

Design and Procedure

We employed a naturalistic stress manipulation design, capitalizing on the university examination period as an ecologically valid acute stressor. The study comprised three assessment waves across 2.5 months:

1. **Baseline assessment** (T1): Administered 3-4 weeks before scheduled exams, participants completed the full PID-5 questionnaire and provided vocal recordings in a laboratory setting.
2. **Pre-exam assessment** (T2): The day before a major course examination, participants recorded vocal samples in the same laboratory.
3. **Post-exam assessment** (T3): The day after the examination, participants provided final vocal recordings.

Between T1 and T3, participants completed twice-weekly ecological momentary assessments (EMA) via smartphone application, yielding an average of 27.0 EMA observations per participant (range: 12-31).

Measures

Personality Pathology. **Full PID-5 (Baseline).** At T1, participants completed the 220-item Personality Inventory for DSM-5 (Krueger et al., 2012), which assesses five maladaptive trait domains: Negative Affectivity (α = [INSERIRE]), Detachment (α = [INSERIRE]), Antagonism (α = [INSERIRE]), Disinhibition (α = [INSERIRE]), and Psychoticism (α = [INSERIRE]). Items are rated on a 0-3 scale (0 = very false/often false, 3 = very true/often true). Domain scores were computed as mean item ratings.

Brief PID-5 for EMA. To reduce participant burden while maintaining construct coverage, we administered a 15-item brief version in EMA assessments (3 items per domain). Items were selected based on factor loadings from a larger sample from the same population (N > 1,000) to maximize domain representativeness while minimizing redundancy.

EMA data were collected via the m-Path smartphone application (RoQua, Tilburg, Netherlands), a validated platform for ambulatory assessment research. Participants received push notifications twice weekly on non-consecutive days between 18:00 and 20:00 over the 2.5-month study period. Each prompt requested ratings of current affect states and brief personality-relevant items. Items were rated on the same 0-3 scale used in the full PID-5, adapted to assess "how you have felt/behaved since the last assessment."

In addition to routine EMA assessments, two exam-related prompts were administered: one immediately before a scheduled exam (same day, 1-4 hours prior) and one the day after exam completion. These exam-linked assessments were paired with voice recordings (see Voice Recordings section) to capture stress-related vocal changes.

Data quality and compliance. Participants with fewer than 50% response rate to EMA prompts were excluded from analysis prior to data processing to ensure adequate sampling of trait-relevant behaviors. Domain scores were computed by averaging items within each domain across all completed assessments, then person-mean centering to create stable trait estimates while removing individual response style effects.

Voice Recordings. At each of the three laboratory sessions (T1, T2, T3), participants recorded sustained vowel phonations, a coarticulation task, and a standardized sentence in a sound-attenuated room. Recordings were made using a Shure SM58 microphone at 44.1 kHz sampling rate, positioned 15 cm from the participant's mouth at a 45 deg. angle to minimize lateral distortions.

Stimuli. Participants produced: 1. Three repetitions of sustained Italian cardinal vowels (/a/, /i/, /u/) for at least 3 seconds each 2. A coarticulation task: counting from 1 to 10 in Italian 3. A standardized constantly-voiced Italian sentence: "Io amo le aiuole della mamma" (English: "I love mother's flowerbeds")

All recordings were performed with conversational pitch and loudness in quiet rooms to maintain ecological validity while ensuring acoustic quality. Participants were instructed to maintain consistent microphone positioning across all sessions.

Recording quality control. Audio files were automatically partitioned into segments corresponding to each vowel repetition, each number, and the standardized sentence, yielding 20 total segments per session. Visual inspection confirmed adequate signal-to-noise ratio and absence of technical artifacts (e.g., clipping, environmental noise intrusions).

Acoustic Feature Extraction. Acoustic features were extracted using two complementary approaches: (1) traditional vocal parameters from sustained vowels, and (2) mel-frequency cepstral coefficients (MFCCs) from continuous speech.

Sustained vowel analysis. The open-source BioVoice software (version [INSERIRE]; [CITATION]) was used to extract 37 acoustic parameters from sustained vowel phonations, spanning frequency and time domains. Key features included:

- **Fundamental frequency (F0).** Mean and median F0 were computed across the sustained portion of each vowel (excluding onset/offset), indexing vocal pitch and laryngeal tension. F0 standard deviation quantified pitch stability. The time instance of maximum F0 (T0) was also extracted as a marker of phonatory dynamics. Higher F0 reflects increased vocal fold tension associated with arousal and stress, representing one of the most robust acoustic markers of altered psychological states (Giddens et al., 2013).

- **Formant frequencies.** Mean F1 (first formant) and F2 (second formant) were extracted for each vowel, along with their minimum and maximum values. Formants index articulatory precision and vocal tract configuration. F2, in particular, reflects tongue positioning and articulatory control, with variability (F2 SD) indicating consistency of articulation.

- **Voice quality parameters.** Jitter (cycle-to-cycle F0 variation, expressed as percentage) and voiced unit duration were extracted to assess glottal stability and phonatory control. Higher jitter values suggest compromised vocal control or irregular vocal fold vibration.

- **Noise-to-harmonics ratio (NNE).** Normalized noise energy (expressed in dB) quantifies aperiodicity in the voice signal, with higher (less negative) values indicating breathy or rough voice quality reflecting incomplete glottal closure or increased tension.

Theoretical rationale. These parameters were selected based on neurophysiological evidence that stress increases general muscular tone, including at the laryngeal level, causing vocal folds to stretch and vibrate more rapidly (F0 increase). Additionally, stress-induced tension in articulators (tongue, jaw, soft palate) alters resonance patterns, reflected in formant shifts and reduced variability (Scherer, 1986; Giddens et al., 2013).

Mel-frequency cepstral coefficients (MFCCs). For the standardized sentence, we computed MFCCs to capture the spectral envelope of continuous speech. MFCCs approximate the human auditory system's nonlinear frequency response and have been extensively validated for emotion and stress recognition (Eyben et al., 2015; Rachman et al., 2018).

The MFCC extraction pipeline was implemented in MATLAB R2023a (The MathWorks, Natick, MA, USA) using the following parameters:

- **Frame-based analysis:** 25ms Hamming windows with 15ms overlap (10ms step)
- **Mel filterbank:** 13 triangular filters spanning 0-8000 Hz
- **Output:** 13 MFCCs per frame (MFCC1-MFCC13)

Each MFCC captures different spectral properties: lower coefficients (MFCC1-MFCC5) reflect broad spectral shape and resonance (related to vocal tract configuration), while higher coefficients (MFCC6-MFCC13) capture fine spectral details including consonantal articulation and aspiration noise.

For each MFCC coefficient across all frames in a sentence, we computed eight summary statistics to characterize distributional properties:

1. Mean (central tendency)
2. Standard deviation (variability/stability)
3. Median (robust central tendency)
4. Interquartile range (IQR; robust variability measure, less sensitive to outliers)
5. Skewness (asymmetry of distribution)
6. Kurtosis (tail heaviness)
7. 25th percentile (lower quartile)
8. 75th percentile (upper quartile)

This yielded 104 MFCC-derived features per sentence (13 coefficients × 8 statistics). Reduced variability (lower SD/IQR) in MFCCs under stress likely reflects constrained articulatory movements due to increased muscular tension, while shifts in mean/median values may indicate altered resonance patterns or prosodic flattening.

Quality control and preprocessing. All acoustic measures were visually inspected using Praat spectrograms and waveform displays. Values beyond 3 SD from the mean were flagged for manual review to identify potential extraction errors (e.g., octave jumps in F0 tracking, formant tracking failures). No systematic outliers requiring exclusion were identified. For analyses, acoustic features were z-score standardized within each feature to enable comparison across parameters with different units and scales.

Data Quality and Missingness

EMA Compliance and Quality Control. Participants with fewer than 5 or more than 40 EMA assessments were excluded prior to analysis (n excluded = [INSERIRE]). We implemented additional quality checks to identify careless responding:

1. **Within-subject variability:** Participants with SD < 0.30 on the Negative Affectivity composite (which should exhibit temporal variation) were flagged for review.
2. **Response patterns:** Excessive use of scale endpoints or preference for round numbers (>80% responses divisible by 10 on 0-100 visual analog scales) indicated potential inattention.
3. **A priori exclusions:** Based on suspicious response patterns identified in preliminary screening, n = [INSERIRE] participants were excluded before analysis.

Final sample included N = 141 participants with complete voice data and valid EMA responses.

Missing Data in Baseline PID-5. Baseline PID-5 data were missing for 93 observations (22% of total) due to participants entering the study through different recruitment streams. Missing values were imputed using random forest imputation (missRanger package; Mayer, 2019) with predictive mean matching (k = 3 nearest neighbors). EMA PID-5 domain scores were included as auxiliary variables given their strong convergent validity (see Results). Sensitivity analyses confirmed that results were robust to imputation approach (complete-case analysis yielded similar parameter estimates; see Supplementary Materials).

Convergent Validity of EMA Measures

To establish construct validity of the brief EMA assessment, we computed correlations between person-level EMA domain scores (aggregated across all assessments) and full baseline PID-5 domains. This addresses the critical question of whether intensive sampling with reduced item coverage captures equivalent trait variance to comprehensive single-session assessment.

Selection of Acoustic Parameters. We focused on two acoustic parameters selected for their theoretical relevance to stress-related vocal changes and their complementary information about distinct physiological mechanisms.

Fundamental frequency (F0) is the most extensively validated acoustic marker of psychological stress, reflecting laryngeal muscle tension and autonomic arousal. Meta-analytic evidence demonstrates reliable F0 elevation under diverse stressors including public speaking, cognitive load, and evaluative threat (Giddens, Barron, Byrd-Craven, Clark, & Winter, 2013; Scherer, 2003), with effect sizes typically in the 3–10 Hz range for naturalistic stressors. Physiologically, F0 increases result from sympathetic nervous system activation increasing cricothyroid and vocalis muscle tension, thereby elevating vocal fold stiffness and vibratory frequency (Titze, 1994). This makes F0 a direct acoustic index of the arousal component of stress responses.

Normalized Noise Energy (NNE) quantifies the ratio of harmonic to inharmonic spectral energy, providing an index of phonatory quality and glottal closure completeness (Kasuya, Ogawa, Mashima, & Ebihara, 1986). Unlike perturbation measures (jitter, shimmer) which primarily capture vocal instability, NNE is sensitive to stress-induced changes in phonatory control and effort. Research indicates that acute stress can either increase noise (via incomplete glottal closure under arousal) or decrease it (via compensatory hyperadduction and pressed phonation), with the direction dependent on individual coping strategies and task demands (Mendoza & Carballo, 1998; Scherer, 2003). By examining NNE alongside F0, we could distinguish arousal-driven pitch changes from control-related adjustments in phonatory quality, testing whether stress induces vocal degradation (increased noise) or compensatory control (reduced noise).

This dual-parameter approach aligns with multidimensional models of vocal stress responses (Giddens et al., 2013) and allows us to test competing hypotheses about how personality traits might selectively moderate arousal versus control components of the stress response. We did not examine other acoustic features (e.g., formant frequencies, spectral tilt, speech rate) as these are more strongly influenced by linguistic content and articulation than by the phonatory physiology most directly affected by autonomic stress responses.

Statistical Analysis

All analyses were conducted in R (version 4.5) using Bayesian multilevel models implemented in brms (Burkner, 2017) with cmdstanr backend. We adopted a Bayesian framework to enable principled uncertainty quantification, incorporate prior knowledge, and estimate complex variance structures without convergence issues common in frequentist multilevel modeling. To evaluate moderation effects, we prioritized direction certainty ($\mathrm{PD} = \max[\mathrm{P}(\ > 0), \mathrm{P}(\ < 0)]$) over magnitude precision, classifying effects with $\mathrm{PD} > 0.95$ as showing strong directional evidence.

Model Specification. **Moderation models.** For each acoustic outcome (6 features × 3 vowels = 18 total), we estimated:

...

Where i indexes participants, j indexes timepoints, Stress and Recovery are orthogonal contrast codes (Stress: baseline = -0.5, pre-exam = 0.5, post-exam = 0; Recovery: baseline = 0, pre-exam = -0.5, post-exam = 0.5), and all trait predictors were person-mean centered. Random effects included random intercepts and uncorrelated random slopes for temporal contrasts.

Likelihood families. We selected likelihood distributions based on outcome properties:

- F0 mean, F2 mean, NNE: Gaussian (unbounded continuous)
- F0 SD, F2 SD, Jitter: Lognormal (positive continuous, right-skewed)

For F2 mean, robust regression with Student-t likelihood was used to accommodate occasional extreme values.

Prior Specification. Priors were weakly informative, centered on realistic parameter ranges while allowing data to dominate inference:

These priors were derived from pilot data and domain knowledge about typical vocal parameter ranges in female speakers.

Estimation. Models were estimated using Hamiltonian Monte Carlo with 4 chains of 5,000 iterations each (2,500 warmup). Convergence was assessed via Rhat < 1.01 and effective sample size > 400. Adaptation parameters (adapt_delta = 0.995, max_treedepth = 18) prevented divergent transitions. For models with convergence difficulties, we increased iterations or simplified random effects structures.

Inference. Effects were considered credible if 95% credible intervals excluded zero. We report posterior means and 95% CIs throughout. For key hypotheses, we computed Bayes factors comparing moderation models to null models without interactions.

Model Comparison: EMA vs. Baseline PID-5. To assess whether intensive EMA sampling provided predictive value beyond traditional baseline assessment, we compared two sets of moderation models:

1. **EMA models:** Using person-aggregated EMA trait scores (as described above)
2. **Baseline models:** Using full baseline PID-5 domain scores

Models were identical in structure, differing only in predictor variables. We compared predictive accuracy using: - **Bayesian R2:** Proportion of variance explained - **Leave-One-Out Cross-Validation (LOO-IC):** Expected out-of-sample predictive accuracy

We expected EMA and baseline measures to show equivalent predictive accuracy if the brief assessment captured core trait variance, or superior accuracy for EMA if intensive sampling reduced measurement error.

Data and Code Availability. All analysis code and de-identified data will be made publicly available upon publication at [OSF LINK]. Models were fit using brms 2.21 with cmdstanr 0.7.

Results

We report results in four sections. First, we establish the main effects of exam-related stress on vocal production, focusing on fundamental frequency (F0) and normalized noise energy (NNE). Second, we examine whether PID-5 personality domains moderate these stress responses. Third, we compare the predictive performance and precision of EMA-based versus baseline PID-5 assessments. Fourth, we evaluate within-person temporal covariation between momentary personality states and acoustic parameters. All analyses used hierarchical Bayesian models implemented in Stan via the rstan package, with four chains of 4,000 iterations each (2,000 warmup). Convergence was verified through R-hat statistics (all < 1.01) and trace plot inspection. We report posterior medians with 95% credible intervals (CrIs) and directional probabilities.

Main Effects of Exam-Related Stress on Vocal Acoustics

We first examined whether acute exam-related stress altered fundamental frequency and glottal noise independently of personality traits. Hierarchical models incorporated random intercepts and random slopes for two orthogonal contrasts: a stress contrast (c_1) comparing pre-exam to baseline recordings, and a recovery contrast (c_2) comparing post-exam to pre-exam recordings. This parameterization allowed us to distinguish the immediate impact of anticipatory stress from subsequent recovery dynamics.

Fundamental frequency. Descriptive statistics revealed a progressive pattern across timepoints. Mean F0 at baseline was 190.7 Hz (SD = 22.0), increasing to 194.0 Hz (SD = 21.9) immediately before the exam and then declining slightly to 192.5 Hz (SD = 23.6) following the exam. The hierarchical model confirmed robust stress-related elevation. The intercept parameter α, representing the estimated F0 at baseline, had a posterior median of 192.48 Hz (MAD = 1.73, 95% CrI [189.07, 195.96]). The stress contrast β_1 showed a clear positive effect: F0 increased by 3.27 Hz (MAD = 1.25, 95% CrI [0.81, 5.71]) when comparing pre-exam to baseline recordings, with $P(_1 > 0)% = 0.995. This finding indicates that acute academic stress reliably elevates vocal pitch, consistent with increased laryngeal tension and autonomic arousal. The recovery contrast β_2 was essentially null (median = 0.14 Hz, MAD = 1.24, 95% CrI [-2.34, 2.59], $P(\beta_2 > 0) = 0.542$), indicating that F0 plateaued after the exam with minimal further change during the brief recovery period.

Between-person variability was substantial. The standard deviation of random intercepts was $\tau_1 = 19.86$ (95% CrI [17.68, 22.44]), reflecting considerable individual differences in baseline vocal pitch. The standard deviation of random slopes for the stress contrast was $\tau_2 = 1.08$ (95% CrI [0.04, 4.45]), indicating modest heterogeneity in stress reactivity after accounting for personality moderation (see below). Residual variability within individuals was $\sigma = 9.10$ Hz (95% CrI [8.34, 9.96]).

Normalized noise energy. In contrast to F0, NNE exhibited a pattern consistent with reduced glottal noise under stress. Descriptively, mean NNE at baseline was -26.55 dB (SD = 2.64), decreasing to -27.09 dB (SD = 3.25) at the pre-exam assessment and remaining relatively stable at -26.98 dB (SD = 2.91) post-exam. More negative NNE values indicate a more periodic, harmonically stable signal. The hierarchical model confirmed systematic stress-induced reduction in glottal noise. The intercept had a posterior median of -26.87 dB (MAD = 0.20, 95% CrI [-27.28, -26.47]). The stress contrast showed a robust negative effect: NNE decreased by 0.79 dB (MAD = 0.31, 95% CrI [-1.30, -0.30]), with $P(\beta_1 < 0) = 0.995$. The recovery contrast was $\beta_2 =$ -0.19 dB (MAD = 0.30, 95% CrI [-0.69, 0.31]). The 95% credible interval includes zero and the directional probability is weak ($P(\beta_2 < 0) = 0.219$), indicating minimal systematic change during the post-exam period. Random effects estimates revealed considerable between-person heterogeneity in baseline NNE ($\tau_1 = 2.14$, 95% CrI [1.85, 2.46]) and in stress-related change ($\tau_2 = 0.71$, 95% CrI [0.06, 1.56]). Residual variability was $\sigma = 1.98$ dB (95% CrI [1.78, 2.17]).

Summary. Exam-related stress produced dissociable changes in vocal production. Fundamental frequency increased robustly under stress, reflecting heightened autonomic arousal and laryngeal tension. In contrast, NNE decreased, indicating reduced glottal noise and a more controlled, periodic phonatory signal. These patterns suggest that acute stress does not simply destabilize the voice but instead induces simultaneous increases in physiological arousal (indexed by F0) and compensatory phonatory control (indexed by reduced noise). The consistent directionality and strong posterior probabilities for the stress effects (P > 0.99 for both F0 and NNE) underscore the reliability of these vocal signatures of stress, though recovery effects showed weaker evidence with credible intervals including zero for both parameters.

Personality Moderation of Vocal Stress Responses

To examine whether PID-5 personality domains moderated vocal stress responses, we added to the previously-described models a trait × contrast interactions. Personality traits were modeled as latent variables derived from EMA assessments, incorporating explicit measurement error correction. We report moderation effects as the change in the stress or recovery contrast effect associated with a one-standard-deviation increase in the trait. We first present F0 moderation results, then examine whether personality domains differentially modulate voice quality (NNE).

Arousal-related pitch responses (F0). For F0, we estimated ten moderation parameters: five domains (Negative Affectivity, Detachment, Antagonism, Disinhibition, Psychoticism) crossed with two contrasts (stress, recovery). Table 1 presents posterior medians, 95% credible intervals, and directional probabilities for each interaction. Among these ten tests, only one showed clear evidence of moderation: Negative Affectivity amplified the stress-induced increase in F0 ($\gamma_1 = 3.14$ Hz per SD, 95% CrI [0.37, 5.89], PD = 0.97). This effect indicates that individuals higher in emotional reactivity and stress sensitivity (Negative Affect) exhibited stronger vocal arousal responses during anticipatory stress. No other stress-phase moderation effects exceeded conventional evidence thresholds (all PD < 0.60).

For the recovery contrast, Antagonism showed the strongest moderation ($\gamma_2 = 3.16$ Hz per SD, 95% CrI [0.51, 5.78], PD = 0.97), suggesting that individuals higher in callousness and interpersonal hostility (Antagonism) exhibited continued F0 elevation during the post-exam period. However, this effect should be interpreted cautiously given the weak main effect of the recovery contrast itself and the relatively small sample for detecting interaction effects in the recovery phase.

Voice Quality Moderation (NNE). In contrast to F0, NNE showed minimal moderation during the stress phase, with all domains exhibiting weak directional certainty (all PD < 0.83 for γ_1). However, the recovery phase revealed a distinct pattern: Psychoticism demonstrated strong directional certainty for recovery moderation ($\gamma_2 = 0.88$ dB, 95% CrI [0.05, 1.72], PD = 0.96, SNR = 1.74). This effect indicates that individuals higher in odd or eccentric thinking (Psychoticism) exhibited less negative NNE values (i.e., increased glottal noise) during the post-exam period, reflecting reduced phonatory control following stress exposure. Antagonism showed suggestive evidence for recovery moderation in the opposite direction ($\gamma_2 =$ -0.43 dB, 95% CrI [-1.37, 0.49], PD = 0.82), though this fell below conventional evidence thresholds. No other domains showed meaningful NNE modulation (Table 2).

The selective moderation patterns reveal striking domain-specificity in personality influences on vocal stress responses. Whereas Negative Affectivity reliably shaped arousal-related pitch responses during stress anticipation, voice quality (NNE) showed a completely distinct pattern: only Psychoticism modulated NNE, and exclusively during the recovery phase ($\gamma_2 = 0.88$ dB, PD = 0.96). This dissociation suggests that different personality domains influence distinct temporal phases and acoustic dimensions of stress responses. Internalizing traits (Negative Affectivity) appear to primarily modulate autonomic arousal indexed by F0 during stress exposure, whereas thought disorder characteristics (Psychoticism) influence phonatory control mechanisms reflected in glottal noise, particularly during stress de-escalation.

[TABLE 1 HERE: PID-5 Domain × Stress/Recovery Interactions for F0]

[TABLE 2 HERE: PID-5 Domain × Stress/Recovery Interactions for NNE]

Comparing EMA-Based and Baseline PID-5 Assessments

A methodological question central to our design was whether repeated measurement via EMA provided advantages over comprehensive single-occasion assessment. To address this, we compared three modeling approaches using leave-one-out cross-validation (LOO-CV): (1) EMA-only, incorporating three EMA assessments within a latent variable measurement model; (2) Baseline-only, using the full 220-item PID-5 from a single administration; and (3) Combined, simultaneously estimating both EMA latent traits and baseline domain scores.

Out-of-sample predictive performance, quantified via expected log pointwise predictive density (ELPD), was comparable across the three approaches (Table 3). The EMA-based model showed numerically the highest ELPD, but differences relative to the Combined model (Δ ELPD = -3.0, SE = 3.6) and Baseline-only model (Δ ELPD = -4.6, SE = 4.0) did not exceed the conventional threshold for meaningful differences ($|\Delta| < 2$ SE). This equivalence indicates that ambulatory assessment and comprehensive single-occasion assessment provide similar predictive accuracy for vocal F0 trajectories during acute stress.

However, examination of moderation effect estimates revealed an important distinction. Table 3 presents a focused comparison for Negative Affectivity, the domain showing the strongest stress moderation. The EMA-based model yielded a more precise estimate ($\gamma_1 = 3.07$ Hz, 95% CrI $[-0.44, 6.55]$, PD = 0.96) compared to the baseline model ($\gamma_1 = 2.65$ Hz, 95% CrI $[-2.20, 7.52]$, PD = 0.86). The EMA estimate showed a 28% narrower credible interval and stronger directional evidence. This pattern was consistent across other PID-5 domains: EMA-derived estimates systematically showed tighter uncertainty bounds despite comparable point estimates (Supplementary Table S2).

The Combined model, which simultaneously estimated both measurement approaches, produced moderation estimates intermediate between the two single-source models. Neither measurement approach dominated when both were included, suggesting that EMA and baseline assessment capture largely overlapping rather than complementary variance in predicting vocal stress reactivity. Together, these results indicate that while EMA does not improve aggregate predictive performance, it does enhance inferential precision for moderation effects—a distinction relevant for theory testing even when forecasting accuracy is equivalent.

Within-Person Temporal Covariation

Finally, we examined whether momentary fluctuations in personality states—assessed via EMA immediately before each voice recording—covaried with concurrent F0 levels. This tests whether trait-level moderation effects have within-person analogs: do individuals show higher F0 when they report elevated negative affectivity in the moment?

We compared hierarchical specifications with fixed effects (population-average slopes only) versus random slopes (allowing individual differences in within-person associations). Random slopes models substantially outperformed fixed-effects specifications ($R^2 = 35\%$ vs 2.5%; ELPD difference ≈ 40 points). However, with only three observations per person, posterior distributions for individual slopes were extremely wide—99.7% of participant $\times$ domain combinations yielded credible intervals spanning zero.

This pattern reveals a critical distinction: the population shows clear evidence of slope heterogeneity (nonzero σ_β for several domains), but we cannot reliably identify which specific individuals have nonzero slopes. This reflects statistical power limitations rather than modeling failure. Denser temporal sampling would be required to resolve individual-level within-person dynamics.

Discussion

This study demonstrates how integrating passive acoustic monitoring with ecological momentary assessment (EMA) reveals context-dependent expression of personality pathology in real-world settings. We investigated whether exam-related stress, as a naturalistic, high-stakes evaluative context, altered vocal production. Furthermore, we investigated whether these acoustic changes were moderated by personality pathology traits measured via ambulatory assessment. Three key findings emerged. Firstly, acute academic stress produced systematic changes in vocal acoustics; fundamental frequency increased by 3.27 Hz, while normalised noise energy decreased by 0.65 dB. This reflects simultaneous physiological arousal and enhanced phonatory control. Secondly, personality pathology dimensions were found to selectively moderate these stress responses: Negative Affectivity was found to amplify acute pitch elevation, while Antagonism was found to prolong post-stressor vocal tension. Thirdly, brief EMA-based personality assessments (15 items assessed repeatedly) were found to be as predictive as comprehensive single-occasion questionnaires (220 items), while providing superior precision for estimating moderation effects. These findings broaden the research on context-sensitive personality assessment by establishing vocal acoustics as an additional, unobtrusive method of capturing real-time psychophysiological responses to naturalistic stressors.

Vocal Acoustics as Passive Indicators of Situational Stress

The robust main effects of exam stress on vocal production show that acoustic features are sensitive to naturalistic psychological stressors, even when controlling for individual differences. The 3.27 Hz elevation in fundamental frequency observed during pre-exam stress is consistent with extensive psychophysiological research linking autonomic arousal to increased laryngeal muscle tension and subglottal pressure. Importantly, this effect emerged in ecologically valid recordings (i.e., reading tasks completed in participants' natural environments) rather than laboratory-induced stress paradigms. This demonstrates that vocal markers can detect stress responses as they unfold in everyday contexts.

The concurrent decrease in normalised noise energy, indicating reduced glottal turbulence and more regular phonation, reveals a more nuanced picture than that of simple stress-induced vocal degradation. Rather than destabilizing voice production, exam stress appears to induce a “controlled tension” state involving heightened physiological activation (as indexed by F0) coupled with compensatory motor control (as indexed by cleaner phonation). This pattern may reflect performance-oriented vocal behavior under evaluative conditions, where speakers unconsciously adopt more effortful, precise articulation to maintain communicative effectiveness despite internal arousal. From a theoretical perspective, this dissociation between arousal-related (F0) and control-related (NNE) acoustic dimensions suggests that vocal stress signatures are multidimensional, with different acoustic features indicating distinct underlying processes.

The minimal recovery effects ($\beta_2 \approx 0$ for F0; β_2 = -0.22 dB with 95% CrI including zero for NNE) indicate that stress-induced vocal changes persisted beyond exam completion. F0 remained approximately 3 Hz above baseline during the post-exam assessment, while NNE showed no clear evidence of normalization. This temporal pattern suggests that acoustic stress markers may recover more slowly than some peripheral physiological indicators (e.g., heart rate), though our sparse sampling design limits firm conclusions about recovery dynamics. Future work employing dense temporal sampling, enabled by smartphone-based voice collection triggered multiple times daily, could characterise individual stress reactivity and recovery trajectories more precisely over longer periods. This temporal pattern extends beyond the typical assessment window and points to a limitation of our sparse-sampling design. While this design effectively captured broad stress phases (baseline, anticipation and the immediate aftermath), it could not resolve the finer dynamics of recovery. Future work employing dense temporal sampling, enabled by smartphone-based voice collection triggered multiple times daily, could characterise individual stress reactivity and recovery trajectories more precisely over longer periods.

Personality Pathology and Context-Dependent Vocal Expression

Against the backdrop of these modest average effects (3.27 Hz for F0 and 0.79 dB for NNE), the magnitude of the personality moderation effects was comparable or larger, demonstrating substantial individual heterogeneity in the acoustic manifestation of stress. This finding directly addresses the special issue's emphasis on person-environment interactions: personality traits do not merely predict fixed behavioural tendencies, but actively influence *how individuals respond to specific situational contexts*.

Negative affectivity, characterised by emotional lability, anxiety and insecurity in separation, reliably amplifies stress-induced pitch elevation. Individuals scoring one standard deviation above the mean for Negative Affectivity exhibited a total F0 increase of approximately 6.4 Hz (3.27 Hz baseline effect + 3.14 Hz moderation), which is almost double the average stress response. This pattern aligns with theoretical models of negative affectivity as a core vulnerability factor for stress reactivity, reflecting heightened sensitivity to threat cues and dysregulated autonomic arousal. Importantly, this moderation was specific to the acute stress phase (pre-exam) and to F0 rather than noise parameters. This suggests that negative affectivity primarily amplifies the arousal-driven component of stress responses rather than destabilising vocal production more broadly.

Antagonism showed the opposite temporal pattern, selectively moderating the **recovery contrast** (post-exam vs. pre-exam) rather than the initial stress reactivity. Individuals high in Antagonism exhibited continued F0 elevation following the exam (2.89 Hz per SD), while others showed normalization. This sustained vocal tension may reflect difficulty disengaging from evaluative contexts, which aligns with the core features of antagonism: hypersensitivity to perceived slights, hostile attribution biases and impaired stress recovery. From a clinical perspective, this finding suggests that different personality pathology dimensions confer vulnerability at different phases of the stress-response cycle: Negative Affectivity increases acute reactivity, while Antagonism impairs subsequent regulation.

The domain-specificity of these moderation effects—Detachment, Disinhibition, and Psychoticism showed minimal reliable moderation—is theoretically informative. Rather than reflecting a general "personality dysregulation" factor, vocal stress reactivity appears to be specifically linked to dimensions involving *emotional sensitivity* (Negative Affectivity) and *interpersonal dysfunction* (Antagonism). This specificity reinforces dimensional models such as the Alternative Model for Personality Disorders (AMPD), which posit that different pathological traits involve distinct mechanisms rather than representing severity along a single continuum.

Critically, NNE and F0 showed **domain-specific and phase-specific modulation patterns**. While Negative Affectivity and Antagonism modulated F0 during stress and recovery respectively, NNE was uniquely modulated by Psychoticism during the recovery phase ($\gamma_2 = 0.88$ dB, PD = 0.96). This triple dissociation—across personality domains, acoustic parameters, and temporal phases—has important theoretical implications.

First, it suggests that vocal stress responses are **multidimensional**, comprising at least two dissociable components: (1) arousal-driven pitch changes (F0) sensitive to internalizing traits during stress exposure, and (2) phonatory control mechanisms (NNE) influenced by thought disorder characteristics during recovery. Second, the Psychoticism effect's specificity to the recovery phase may reflect disrupted self-monitoring or altered interoceptive awareness characteristic of this domain. While most individuals maintain compensatory vocal control after stress (as indicated by minimal NNE recovery main effects), high-psychoticism individuals show increased glottal noise, possibly reflecting reduced attention to phonatory precision or slower physiological de-escalation.

This domain-specificity has practical implications for voice-based digital phenotyping. F0 appears most informative for detecting internalizing pathology and acute stress reactivity, whereas NNE may index thought disorder symptoms and recovery processes. Multivariate approaches incorporating both dimensions could enhance phenotyping precision by capturing complementary aspects of personality-stress interactions.

Methodological Innovation: Multimodal Ambulatory Assessment

A central contribution of this study is demonstrating the value of **integrating passive sensing (voice acoustics) with active self-report (EMA)** within a unified ambulatory assessment framework. This multimodal approach directly addresses the special issue's call for methods that "go beyond self-report to enhance ecological validity" and "capture dimensions of context…not easily accessible through EMA questionnaires alone."

Voice recordings offer several advantages as a complement to traditional EMA:

1. Reduced participant burden. Voice samples require minimal active engagement (30-second reading tasks) compared to extended questionnaires, potentially increasing compliance in intensive longitudinal designs. Future implementations could leverage entirely passive collection during natural phone conversations or voice memos, eliminating participant burden entirely.

2. Objective psychophysiological data. Unlike self-reported stress or mood—which are subject to awareness limitations, recall biases, and social desirability—acoustic features reflect automatic physiological processes (laryngeal tension, respiratory patterns) that operate below conscious control. This provides convergent validation of psychological states through a different measurement channel.

3. Temporal resolution. Voice can be sampled at much higher frequencies than practical for questionnaire-based EMA. Smartphones could theoretically analyze every phone call, voice message, or interaction with voice assistants, providing dense time-series data on stress dynamics. Our sparse three-timepoint design represents a conservative implementation; continuous or event-triggered voice monitoring could reveal within-day fluctuations and acute stress episodes missed by scheduled assessments.

4. Ecological validity. Voice samples collected via smartphone apps capture vocal behavior in participants' natural environments and daily routines, rather than laboratory-constrained speech tasks. This enhances generalizability to real-world contexts where personality pathology actually manifests.

The comparable predictive accuracy of EMA versus baseline personality assessment (Section: Comparing EMA-Based and Baseline PID-5 Assessments) demonstrates that **intensive repeated measurement of brief personality indicators captures equivalent trait variance to comprehensive single-occasion questionnaires**. This finding has important implications for ambulatory assessment design: researchers need not administer extensive trait measures repeatedly when brief indicators assessed across multiple occasions provide comparable information. The 29% narrower credible intervals for EMA-derived moderation estimates reflect enhanced precision through explicit modeling of measurement error across occasions—a key advantage when testing specific theoretical hypotheses about personality-context interactions.

However, the latent variable approach we employed is critical for realizing this advantage. Simply averaging repeated EMA assessments would lose precision by failing to separate true score variance from occasion-specific fluctuations. The measurement model framework treats each EMA occasion as a fallible indicator of an underlying latent trait, accumulating information across assessments to estimate stable individual differences. This methodological refinement is essential for ambulatory personality research and represents a concrete innovation aligned with the special issue's goals.

Temporal Dynamics and the Challenge of Sparse Sampling

Our analysis of within-person temporal covariation revealed a critical limitation: with only three measurement occasions per person, we could not reliably detect individual-level associations between momentary personality states and concurrent vocal acoustics. This finding underscores a fundamental tension in ambulatory assessment design between **temporal density** (many observations per person) and **participant burden** (feasibility of intensive protocols).

The random slopes models indicated substantial population-level heterogeneity in within-person associations (_   2.9 Hz for Negative Affectivity), suggesting that individuals differ meaningfully in how their momentary affective states couple with vocal production. However, individual-level estimates were too imprecise to distinguish from zero (99.7% of credible intervals spanning zero), despite this population-level variability. This distinction is important: **the effect exists at the population level but cannot be reliably detected at the individual level** with current sampling density.

Simulation studies suggest that reliably estimating random slope variances requires 10-20 observations per person—substantially more than our three-timepoint design. Future research implementing this multimodal approach should prioritize denser temporal sampling, potentially through:

- **Event-contingent assessment:** Triggering voice recordings and EMA following identified stressors (detected via passive sensing, calendar events, or self-initiation)
- **Burst designs:** Intensive sampling during high-risk periods (exam weeks, interpersonal conflicts) interspersed with low-frequency baseline monitoring
- **Fully passive collection:** Continuous acoustic monitoring of natural speech (with appropriate privacy protections), eliminating the need for scheduled assessments

The temporal resolution question also relates to conceptual models of context. Our study examined stress as a relatively extended state (exam anticipation spanning hours to days), but personality pathology may be particularly evident in **momentary, dynamic responses** to acute interpersonal triggers—arguments, perceived rejections, criticism—that unfold over minutes rather than days. Capturing these microtemporal processes requires assessment strategies capable of detecting rapid within-person fluctuations, which our design could not accommodate.

Implications for Context-Sensitive Models of Personality Pathology

The selective moderation patterns we observed have important implications for how we conceptualize personality pathology as context-dependent. Rather than treating traits as static dispositions that uniformly amplify or dampen stress responses, our findings suggest **domain-specific, phase-specific moderation**: Negative Affectivity shapes acute reactivity, Antagonism shapes recovery, and other dimensions show minimal vocal stress coupling. This specificity supports transactional models of personality-environment interaction, where different trait facets become activated in different situational contexts.

From a clinical perspective, these findings point toward **personalized stress phenotypes** that could inform intervention targets. Individuals high in Negative Affectivity might benefit from interventions targeting acute stress reactivity (e.g., cognitive restructuring of threat appraisals, autonomic regulation training), while those high in Antagonism might require support for post-stressor recovery and disengagement (e.g., emotion regulation strategies, mindfulness-based approaches). Voice-based ambulatory monitoring could potentially track treatment response by detecting changes in stress-related acoustic signatures over time.

The integration of vocal acoustics with EMA also addresses the special issue's emphasis on capturing "proximal triggers…in ways that illuminate how contextual factors evoke, amplify, or attenuate maladaptive trait expression." By directly measuring psychophysiological responses (voice) alongside self-reported internal states (EMA) within the same naturalistic stressor context, we can characterize the real-time coupling between personality traits, environmental demands, and embodied stress responses. This represents a concrete implementation of multilevel, multimodal assessment that moves beyond generic trait-outcome associations toward mechanistic understanding of person-situation transactions.

Limitations and Future Directions

Several limitations qualify our conclusions and point toward future research directions aligned with the special issue's agenda:

1. Proximal context assessment. While we successfully embedded assessment within a naturalistic stressor (academic examinations), we did not measure fine-grained situational features that might further moderate stress responses. For example, exam difficulty, perceived preparedness, social comparison processes, or prior exam outcomes could all influence stress intensity and recovery. Future research should integrate momentary assessments of **perceived situational characteristics** alongside objective stressors, potentially using experience sampling methods to capture participants' subjective construals of exam contexts.

2. Limited acoustic feature space. We focused on fundamental frequency and glottal noise as theory-driven markers of arousal and vocal quality. However, stress may manifest in spectral characteristics (e.g., spectral tilt, harmonic structure), temporal patterns (speech rate, articulation rate, pause duration), prosodic contours (pitch variability, intonation), or voice quality dimensions (breathiness, roughness, strain) not examined here. Comprehensive acoustic profiling using machine learning approaches could identify additional personality × stress signatures not predicted by existing theory.

3. Lack of physiological validation. The interpretation of F0 changes as reflecting autonomic arousal relies on established psychophysiological theory but was not directly validated in our data. Integrating wearable sensors to capture heart rate variability, electrodermal activity, or cortisol (via salivary sampling) would provide convergent evidence and enable examination of whether vocal changes mediate, moderate, or operate independently of peripheral physiological stress responses.

4. Single stressor type. Academic examinations represent achievement-oriented, evaluative stress contexts but may not generalize to interpersonal stressors (conflict, rejection, criticism) particularly relevant to personality pathology. Given that interpersonal dysfunction is central to most personality disorder conceptualizations, examining vocal responses to social stressors—potentially captured through analysis of natural conversations or conflict discussions—represents a high priority for future research.

5. Sample characteristics. University students experiencing normative academic stress differ from clinical populations with diagnosed personality disorders in both trait severity and stress exposure. The moderation effects we observed may be attenuated in subclinical samples; clinical populations might show larger personality $\times$ stress interactions or different patterns of acoustic reactivity. Replication in treatment-seeking samples with elevated personality pathology is essential for establishing clinical utility.

6. Temporal generalizability. The three-timepoint design captured stress anticipation and immediate aftermath but not longer-term recovery or repeated stress exposure. Chronic stress contexts (e.g., ongoing relationship conflict, academic probation) or repeated stress assessments (multiple exams across a semester) could reveal cumulative effects, sensitization, or habituation processes not detectable in single-episode designs.

7. Mechanism specificity. Acoustic changes could reflect multiple pathways: (a) automatic physiological arousal, (b) strategic emotional regulation (suppressing or expressing distress vocally), (c) cognitive load effects on speech production, or (d) communicative signaling of stress to interaction partners. Disentangling these mechanisms requires experimental manipulations or mediation analyses integrating self-report, physiological, and acoustic data within person-centered analytic frameworks.

8. Predictive validity. While we demonstrate that personality moderates stress-related acoustic changes, we do not know whether these patterns predict clinically meaningful outcomes—psychopathology onset, functional impairment, treatment response, or real-world adaptive behavior. Longitudinal research linking individual differences in vocal stress reactivity to prospective outcomes would establish the clinical significance of these acoustic signatures.

9. Moderation effects. The modest effect magnitudes (3-4 Hz for F0, <1 dB for NNE) and moderate signal-to-noise ratios (SNR 1.5-2.0) indicate personality traits explain small portions of stress-response variance at the population level. While directional certainty was high (PD > 0.95), credible intervals remained wide, limiting individual-level prediction precision.

Future Directions: Advancing Multimodal Ambulatory Assessment

Building on our findings, we propose several directions for advancing context-sensitive personality pathology assessment:

1. Real-time adaptive sampling. Rather than fixed-schedule assessments, future studies could implement algorithms that detect acoustic stress signatures in real time and trigger targeted EMA prompts only when stress episodes are detected. This would reduce participant burden while ensuring dense sampling during clinically relevant moments.

2. Social interaction analysis. Our reading task paradigm isolated individual stress responses but did not capture the interpersonal contexts where personality pathology is most evident. Analyzing natural conversations—using speaker diarization to separate participants' voices and natural language processing to code interaction content—could reveal how personality traits shape vocal behavior in dyadic stress contexts (arguments, support-seeking, conflict resolution).

3. Contextual enrichment through passive sensing. Integrating voice with GPS (location tracking), accelerometry (physical activity), smartphone usage patterns (social media, communication frequency), and calendar data (scheduled stressors) could provide rich contextual information about the situations in which stress responses occur. Machine learning models could then identify situational features that evoke or buffer personality-stress coupling.

4. Intervention applications. Voice-based ambulatory monitoring could support just-in-time adaptive interventions (JITAIs) by detecting stress episodes and delivering personalized coping strategies via smartphone. For example, detecting elevated F0 in someone high in Negative Affectivity could trigger prompts for grounding exercises or cognitive reappraisal.

5. Cross-cultural validation. Voice production varies across languages and cultures in fundamental ways (prosodic norms, expressive display rules, acoustic baselines). Validating stress-related acoustic signatures across diverse linguistic and cultural contexts is essential for generalizability, particularly given the special issue's emphasis on "broader cultural and structural forces" shaping personality expression.

6. Multivariate pattern recognition. Rather than examining individual acoustic features in isolation, machine learning approaches (e.g., support vector machines, random forests, deep learning) could identify complex multivariate acoustic patterns that distinguish stress states or personality profiles. These models might capture subtle feature interactions invisible to univariate analyses.

Conclusion

This study demonstrates that integrating passive acoustic monitoring with intensive ecological momentary assessment provides a powerful multimodal framework for studying context-dependent expression of personality pathology. Vocal acoustics offer an objective, unobtrusive window into real-time psychophysiological stress responses as they unfold in naturalistic environments. By showing that brief EMA-based personality assessment captures equivalent information to comprehensive questionnaires while enabling temporal dynamics analysis, we validate intensive longitudinal approaches for personality research. The selective moderation patterns—Negative Affectivity amplifying acute arousal, Antagonism prolonging recovery—establish that different personality pathology dimensions confer vulnerability at different phases of the stress-response cycle, supporting nuanced, domain-specific models of person-situation transactions.

As voice-based digital phenotyping technologies advance, integrating dimensional personality assessment into acoustic stress models will be essential for developing personalized, context-sensitive mental health monitoring systems. Our findings establish empirical foundations and methodological templates for such work, demonstrating that personality traits measured in daily life shape how stress becomes acoustically manifest in human speech. The modest effect sizes we observe underscore that clinically useful systems will require multimodal data integration, within-person calibration, and explicit modeling of individual differences—but also demonstrate that these requirements are achievable through thoughtfully designed ambulatory assessment protocols.

More broadly, this research illustrates how innovations in ambulatory methodology—combining active self-report, passive sensing, and advanced statistical modeling—can move personality pathology research beyond decontextualized trait-outcome correlations toward mechanistic understanding of how maladaptive patterns emerge from dynamic transactions between individuals and the environments they inhabit. By capturing both the “person” (dimensional traits via EMA) and the “situation” (naturalistic stressors, psychophysiological responses via voice), we advance toward the contextualized, ecologically valid science of personality functioning that this special issue seeks to promote.

References

- Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology, 69*(5), 890.
- Fleeson, W. (2001). Toward a structure-and process-integrated view of personality. *Journal of Personality and Social Psychology, 80*(6), 1011.
- 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, 27*(3), 390-e21.
- Hopwood, C. J., Bleidorn, W., & Wright, A. G. (2022). Connecting theory to methods in longitudinal research. *Perspectives on Psychological Science, 17*(4), 884-894.
- Kent, R. D., & Kim, Y. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics, 17*(6), 427-445.
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*(9), 1879-1890.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality. *Psychological Review, 102*(2), 246.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1-2), 227-256.
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2013). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433-456). Oxford University Press.
- Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology, 129*(1), 56.
- Wright, A. G., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology, 125*(5), 641.

- Wright, A. G., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person. *Assessment, 26*(3), 403-419.

Giddens, C. L., Barron, K. R., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Progressive vocal stress modeling. *Behavioral Sciences*, *3*(4), 571–587.

Kasuya, H., Ogawa, S., Mashima, K., & Ebihara, S. (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. *The Journal of the Acoustical Society of America*, *80*(5), 1329–1334.

Mendoza, E., & Carballo, G. (1998). Acoustic analysis of induced vocal stress by means of cognitive workload tasks. *Journal of Voice*, *12*(3), 263–273.

Scherer, K. R. (2003). Vocal expression of emotion. *Handbook of Affective Sciences*, 433–456.

Titze, I. R. (1994). *Principles of voice production*. Prentice Hall.

Table 1

Personality Moderation of Fundamental Frequency (F0)

Domain	Stress (γ_1)		Recovery (γ_2)	
	Median [95% CrI]	PD	Median [95% CrI]	PD
Negative Affectivity	**3.14 [0.37, 5.89]**	**0.97**	-0.31 [-3.13, 2.51]	0.60
Detachment	-0.38 [-3.10, 2.31]	0.59	-2.02 [-4.84, 0.76]	0.88
Antagonism	0.09 [-2.51, 2.69]	0.52	**3.16 [0.51, 5.78]**	**0.97**
Disinhibition	0.61 [-2.60, 3.82]	0.64	0.68 [-2.51, 3.87]	0.67
Psychoticism	-0.13 [-2.74, 2.46]	0.54	-1.22 [-3.81, 1.32]	0.80

Note. Moderation effects in Hz per SD of trait. PD = Probability of Direction. Bold = strong certainty (PD > 0.95). CrI = Credible Interval.

Table 2

Personality Moderation of Normalized Noise Energy (NNE)

Domain	Stress Moderation (γ_1)		Recovery Moderation (γ_2)	
	Median [95% CrI]	PD	Median [95% CrI]	PD
Negative Affectivity	-0.46 [-1.45, 0.52]	0.83	-0.39 [-1.38, 0.62]	0.79
Detachment	0.29 [-0.69, 1.30]	0.72	0.21 [-0.77, 1.21]	0.66
Antagonism	-0.01 [-0.94, 0.90]	0.51	-0.43 [-1.37, 0.49]	0.82
Disinhibition	0.33 [-0.86, 1.51]	0.71	-0.39 [-1.56, 0.79]	0.74
Psychoticism	-0.02 [-1.04, 0.97]	0.52	**0.88 [0.05, 1.72]**	**0.96**

Note. Moderation effects represent the change in NNE (dB) associated with a one-standard-deviation increase in the trait. Positive values indicate less negative NNE (increased glottal noise). PD = Probability of Direction (maximum of $P(\gamma > 0)$ and $P(\gamma < 0)$). Bold indicates strong directional certainty (PD > 0.95). CrI = Credible Interval.

Table 3

Model Comparison: Out-of-Sample Predictive Performance

	ELPD		LOOIC	
Model	Estimate (SE)	Δ (SE)	Estimate (SE)	p_loo (SE)
EMA	-1247.3 (16.8)	—	2494.6 (33.7)	110.3 (9.2)
Combined	-1250.3 (16.7)	-3.0 (3.6)	2500.6 (33.3)	118.0 (9.4)
Baseline	-1251.9 (16.9)	-4.6 (4.0)	2503.8 (33.8)	109.0 (8.9)

Note. ELPD = Expected Log Predictive Density, LOOIC = Leave-One-Out Information Criterion, SE = Standard Error. The Δ ELPD column shows the difference relative to the best-performing model (EMA).

Table 4

Comparison of Negative Affectivity × Stress Moderation Estimates Across Measurement Approaches

	Negative Affectivity × Stress (γ_1)			
Model	Mean (Hz)	90% CrI	PD	Improvement
EMA	3.07	$[-0.44, 6.55]$	0.96	+28%
Baseline	2.65	$[-2.20, 7.52]$	0.86	–

Note. γ_1 = moderation effect of Negative Affectivity on stress-induced F0 change; PD = probability of direction (proportion of posterior above/below zero); CrI Width = credible interval width; Improvement = precision gain of EMA relative to Baseline [(Baseline width - EMA width) / Baseline width]. The EMA-based estimate shows **28% narrower uncertainty bounds** while maintaining comparable point estimates, reflecting enhanced precision through explicit measurement error modeling across repeated assessments.