Quantitative Voice Analysis Reveals Context-Dependent Expression of Personality Pathology: A Bayesian Multilevel Ambulatory Assessment Study

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

This study examined whether vocal acoustics can capture context-dependent expression of personality pathology during naturalistic stress. Female university students (N = 119) provided voice recordings at baseline, immediately before, and after a course examination, while personality pathology traits (PID-5) were assessed via ecological momentary assessment (EMA) over 2.5 months. Exam stress produced dissociable vocal changes: fundamental frequency (F0) increased by 3.27 Hz, indicating heightened autonomic arousal, while normalized noise energy (NNE) decreased by 0.79 dB, suggesting compensatory phonatory control. Personality traits showed domain- and parameter-specific moderation: Negative Affectivity amplified stress-induced F0 increases during anticipatory stress, Antagonism was associated with sustained F0 elevation during recovery, and Psychoticism uniquely modulated voice quality (NNE) during the post-stressor period. Detachment and Disinhibition showed minimal moderation, consistent with theoretical predictions distinguishing arousal sensitivity from behavioral control. Methodologically, brief EMA-based personality assessment matched comprehensive baseline questionnaires in predictive accuracy while yielding more precise moderation estimates. These findings support transactional models of personality pathology and indicate that multimodal ambulatory assessment, combining passive acoustic sensing with intensive self-report, can capture how maladaptive traits shape real-world stress responses.

Keywords: keywords

Word count: X

Quantitative Voice Analysis Reveals Context-Dependent Expression of Personality Pathology: A Bayesian Multilevel Ambulatory Assessment Study

Introduction

Personality pathology is increasingly conceptualized as a context-sensitive phenomenon, wherein maladaptive traits are not expressed uniformly but are modulated—amplified, attenuated, or qualitatively transformed—by situational demands (Hopwood et al., 2022; Wright & Simms, 2016). Contemporary personality theories converge on the idea that stable individual differences are revealed not through static behavior, but through systematic patterns of response to situational cues. For instance, Mischel and Shoda's (1995) *Cognitive-Affective Personality System* frames personality as a stable set of situation-contingent "if–then" behavioral signatures, reflecting consistent individual differences in reactions to specific psychological features of situations. Similarly, Fleeson's (2001) *Whole Trait Theory* conceptualizes traits as density distributions of momentary states, distinguishing between stable mean levels and dynamic, situation-contingent variability stemming from individual differences in contextual responsiveness. Applied to personality pathology, these frameworks suggest that maladaptive traits not only predict outcomes but also moderate the influence of environmental stressors, thereby shaping domain-specific physiological and behavioral responses.

Despite this strong theoretical consensus, empirical methods for capturing person–situation dynamics in naturalistic contexts remain limited. Although the field broadly agrees that "context matters," research on personality pathology has struggled to operationalize contextual processes as they unfold in daily life (Wright et al., 2019). Traditional assessment approaches conceptualize traits as decontextualized dispositions measured at single time points, largely ignoring the temporal dynamics and situational contingencies through which pathology is expressed. This mismatch between theory, which

emphasizes person-environment interactions, and methodology, which relies on static, decontextualized measurement, limits 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 gap by enabling repeated measurement of states and behaviors in real-world contexts, yielding important insights into within-person variability and situational reactivity (Colpizzi, Trull, Sica, Haney, & Caudek, 2025; Trull & Ebner-Priemer, 2020). However, EMA remains fundamentally constrained by its reliance on self-report. Individuals may lack access to their own physiological stress responses, exhibit biased reporting, or show assessment reactivity, whereby the act of self-monitoring alters the phenomena under study (Barta et al., 2012). Moreover, repeated self-report imposes participant burden that limits sampling density, reducing temporal resolution precisely when it is most needed, that is, during acute stress episodes unfolding over minutes to hours. These limitations highlight the need for objective, unobtrusive, and continuous indicators of stress reactivity that do not depend on introspection.

Acoustic features of voice and speech can reflect autonomic arousal, emotional state, and aspects of motor control that are only partly accessible to conscious awareness (Scherer, 2003). Vocal production involves the coordinated activity of multiple physiological systems across the body, including central and peripheral nervous system processes, respiratory control, and laryngeal motor function. As a result, variations in vocal acoustics provide an indirect window into psychophysiological states as they unfold in real time. Fundamental frequency (F0), the acoustic correlate of perceived pitch, increases reliably under stress due to sympathetic activation that tightens the vocal folds via increased laryngeal muscle tension (Giddens et al., 2013). Voice quality parameters such as normalized noise energy (NNE; indexing glottal noise) and jitter (cycle-to-cycle frequency variability) capture aspects of phonatory stability and vocal control, and are sensitive to

both acute emotional arousal and chronic psychological states (Scherer et al., 2013). Crucially, these features can be extracted from brief, naturalistic speech samples, making voice acoustics a scalable, low-burden tool for ambulatory psychophysiological assessment. Table 1 summarizes the definition, physiological basis, and psychological significance of the three principal acoustic features discussed in this literature.

The relevance of vocal stress reactivity for personality pathology is grounded in transactional models of person–situation interaction (Dietrich & Abbott, 2012). These models posit that stable individual differences reflect differences in stress sensitivity, such as heightened threat reactivity, impaired recovery, or dysregulated arousal, such that acute stressors elicit disproportionate physiological responses in vulnerable individuals (Bolger & Zuckerman, 1995). If personality pathology dimensions indeed index such vulnerabilities, they should systematically shape physiological stress responses in everyday life.

The *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012) operationalizes personality pathology across five broad domains representing maladaptive variants of the Five Factor Model (Widiger & Crego, 2019). Given established links between Big Five traits and physiological stress reactivity (Bibbey, Carroll, Roseboom, Phillips, & Rooij, 2013; Luo, Zhang, Cao, & Roberts, 2023), each PID-5 domain carries theoretical implications for stress-related mechanisms. Negative Affectivity, the maladaptive counterpart of Neuroticism, reflects emotional lability and anxiousness—features associated with altered, though not uniformly heightened, autonomic and neuroendocrine stress responses (Lahey, 2009). Detachment, the maladaptive variant of low Extraversion, captures social withdrawal and anhedonia, with competing implications for stress reactivity: reduced social buffering may impair recovery, while blunted emotional expression could attenuate acute responses. Antagonism, characterized by callousness and interpersonal dominance at the low pole of Agreeableness, may either attenuate stress reactivity through reduced empathic engagement or prolong physiological arousal through conflictual social

interactions. Psychoticism involves cognitive and perceptual dysregulation and may manifest as poorly regulated physiological output even under baseline conditions. Finally, Disinhibition, the pathological extension of low Conscientiousness, is characterized by impulsivity and poor self-regulation, potentially compromising motor planning and behavioral control under stress. Critically, these domain-specific implications concern general physiological stress mechanisms (e.g., autonomic activation, neuroendocrine regulation, motor control); whether and how they manifest in the specific channel of vocal production remains an open empirical question that the present study addresses directly.

Despite these theoretically grounded predictions, empirical evidence on how personality pathology moderates physiological stress responses remains sparse, particularly outside the laboratory. Experimental stress paradigms offer strong internal validity but limited ecological relevance, as artificial stressors may not engage the same psychological processes as real-world evaluative threats. Conversely, field studies of daily stress rely almost exclusively on self-report, reintroducing the limitations noted above. What is needed is an approach that integrates (a) ecologically valid stressors with real consequences, (b) objective physiological, or physiology-linked, indicators independent of self-report, and (c) longitudinal designs capable of modeling both within-person dynamics and between-person moderation.

The present study addresses this need through a multimodal ambulatory assessment design centered on vocal stress markers. We leveraged the university examination period as a naturalistic, time-limited stressor with genuine stakes for participants, collecting vocal samples at baseline, immediately before the exam, and following exam completion. Personality pathology traits were assessed via EMA across a 2.5-month period. This design advances the field in three key ways: it operationalizes stress using a proximal real-world stressor rather than retrospective reports; it employs passive voice-based acoustic sensing to capture objective, physiology-linked vocal indicators; and it enables domain-specific

tests of how personality pathology moderates both acute stress reactivity and post-stressor recovery.

Our analyses were guided by two primary questions: (1) whether vocal parameters exhibit reliable stress-related changes in this naturalistic context, and (2) whether PID-5 personality pathology domains moderate these changes. For the first question, we hypothesized that the examination period would be associated with increased F0, consistent with extensive evidence linking acute stress to elevated pitch via sympathetic activation (Giddens et al., 2013; Scherer, 2003). For voice quality parameters (NNE), we anticipated stress-related degradation reflecting reduced phonatory control, although prior findings suggest that voice quality parameters show smaller and less consistent stress effects than F0 (Giddens et al., 2013; Van Puyvelde, Neyt, McGlone, and Pattyn (2018)).

Regarding moderation by personality pathology traits, we adopted a largely exploratory approach. The general implications for stress physiology outlined above—differential autonomic reactivity, impaired recovery, dysregulated motor output—generate plausible predictions at the level of vocal acoustics, but translating them into directional hypotheses requires evidence that personality traits modulate vocal production under stress specifically. Such evidence is scarce: although personality traits are known to shape physiological stress responses such as cortisol reactivity and heart rate variability (Luo et al., 2023), and emerging work indicates that these differences extend to vocal production under stress (Dietrich & Abbott, 2012), the specific moderating role of PID-5 domains on vocal stress markers has not been examined. The exception was Negative Affectivity, for which we predicted amplified stress-induced F0 increases. This prediction follows directly from the construct's emphasis on threat sensitivity and emotional reactivity, features closely tied to sympathetic overactivation (Lahey, 2009; Ormel et al., 2013), combined with F0's well-established sensitivity to autonomic arousal. For the remaining domains, competing theoretical predictions and inconsistent findings in

adjacent literatures led us to treat moderation effects as empirical questions. For instance, Detachment has been linked to affective inertia and impaired affect repair (Kuppens et al., 2010), but restricted emotional expression could alternatively attenuate vocal stress responses. Disinhibition, defined by impulsivity and poor behavioral control rather than emotional reactivity, shows weak associations with physiological stress markers despite moderating subjective stress perception (Luo et al., 2023), suggesting it may not influence the automatic autonomic mechanisms that drive F0 changes.

Finally, we examined whether intensive EMA-based assessment of personality pathology could predict vocal stress responses as effectively as a comprehensive baseline assessment using the complete PID-5 questionnaire. Rather than contrasting different instruments, this comparison focuses on alternative measurement strategies within the same trait framework—brief, repeated assessments in daily life versus a full-length, cross-sectional assessment. If intensive longitudinal measurement can recover stable trait variance through explicit modeling of measurement error and occasion-specific fluctuations, EMA-based approaches may offer a more efficient and scalable alternative to traditional questionnaire administration, an issue of increasing importance in high-burden ambulatory research designs.

By integrating passive voice-based acoustic sensing with intensive longitudinal assessment in a real-world stress context, this study advances both theory and methodology in personality pathology research. It provides an objective, low-burden index of stress reactivity, tests core predictions of transactional models in daily life, and moves the field beyond static, decontextualized trait assessment toward a dynamic, situated understanding of maladaptive personality processes.

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 (Gelfer & Mikos, 2005), 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 #05/23052025).

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.

The course examination was administered digitally via the Moodle platform, with automated scoring that provided participants with their grade immediately upon completion. This procedural feature ensured that, by the time of the T3 recording, uncertainty about the exam outcome had been resolved.

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). PID-5 domain scale scores were calculated as an average of the three most representative facet scores included under each domain as per formal scoring procedures (American Psychiatric Association, 2014).

Brief PID-5 for EMA. To reduce participant burden while maintaining construct coverage, we administered a brief version of the PID-5 in the EMA assessments, comprising a small subset of items per domain. Items were selected based on their factor loadings and domain representativeness in a large independent sample drawn from the same population, as reported in prior validation studies (Bottesi et al., 2024; Sica et al., 2024). This selection

strategy aimed to maximize construct validity while minimizing redundancy and assessment burden in intensive longitudinal designs.

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, but were rephrased to assess current states (e.g., 'At this moment, I worry a lot about being alone'), whereas the original PID-5 items assess general tendencies (e.g., 'I worry a lot about being alone').

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.

Voice Recordings. At each of the three assessment sessions (T1, T2, and T3), participants produced a set of standardized voice recordings in a sound-attenuated room. The recording protocol included sustained vowel phonations, a coarticulation task, and a standardized sentence. In addition, using the m-Path application, participants recorded a short spontaneous audio message spoken with conversational pitch, loudness, and rhythm, based on predefined prompts described below (Manfredi et al., 2017).

All recordings were obtained in quiet indoor environments. Participants were instructed to position the recording device approximately 15 cm from the mouth at a 45°

angle to minimize plosive artifacts and lateral acoustic distortions. Instructions emphasized maintaining consistent microphone placement across sessions.

Stimuli. Participants were asked to produce the following vocalizations:

1. Three repetitions of sustained Italian cardinal vowels (/a/, /i/, /u/), each held for a minimum of 3 s.
2. A coarticulation task consisting of counting aloud from 1 to 10 in Italian.
3. A standardized, continuously voiced Italian sentence: *"Io amo le aiuole della mamma"* (English translation: *"I love my mother's flowerbeds"*).

All speech tasks were performed using conversational pitch and loudness to preserve ecological validity while ensuring sufficient acoustic quality for analysis.

For sustained-vowel features, acoustic parameters were extracted separately for each vowel token (/a/, /i/, /u/; three repetitions each) and then averaged across vowels and repetitions within each session to yield a single session-level estimate per participant.

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

Sustained vowel analysis. The open-source BioVoice software (Morelli & Manfredi, 2019) was used to extract frequency- and time-domain acoustic parameters from sustained vowel productions. Analyses focused on the steady-state portion of each vowel, excluding onset and offset segments to ensure stable phonatory conditions. For each recording session, measures were computed separately for each vowel token (/a/, /i/, /u/) and then averaged across vowels and repetitions.

Fundamental frequency (F0). Mean and median F0 were calculated as summary

indices of vocal fold vibratory rate during sustained phonation, while the standard deviation of F0 indexed short-term pitch variability. In addition, the temporal location of the maximum F0 value (T0) was extracted as a marker of phonatory dynamics. Fundamental frequency is the most extensively validated acoustic correlate of psychological stress and arousal, reflecting changes in laryngeal muscle tension and autonomic activation. Meta-analytic and experimental evidence consistently demonstrates reliable increases in F0 under a wide range of stressors, including public speaking, cognitive load, and evaluative threat (Giddens, Barron, Byrd-Craven, Clark, & Winter, 2013; Scherer, 2003). Physiologically, these increases are attributed to sympathetic nervous system activation, which enhances cricothyroid and vocalis muscle tension, increases vocal fold stiffness, and consequently elevates vibratory frequency (Titze, 1994). As such, F0 provides a direct acoustic index of the arousal component of the stress response.

Normalized Noise Energy (NNE). Normalized noise energy, expressed in decibels, quantifies the relative contribution of aperiodic (noise) energy to the overall voice signal and serves as an index of phonatory quality and glottal closure completeness (Kasuya, Ogawa, Mashima, & Ebihara, 1986). Higher NNE values indicate a greater proportion of noise relative to harmonic energy and are typically associated with reduced or incomplete glottal closure and breathier voice quality. Unlike perturbation measures such as jitter and shimmer, which primarily capture cycle-to-cycle instability, NNE is sensitive to broader stress-related changes in phonatory control and vocal effort. Empirical findings suggest that acute stress can have bidirectional effects on vocal noise: it may increase NNE through incomplete glottal closure and increased airflow turbulence under heightened arousal, or decrease NNE through compensatory hyperadduction and pressed phonation aimed at maintaining vocal control (Mendoza & Carballo, 1998; Scherer, 2003). The direction of these effects appears to depend on individual coping strategies and task demands. Accordingly, NNE may capture individual differences in stress-related vocal regulation, reflecting whether arousal is accompanied by vocal degradation or

compensatory control. Examining NNE alongside F0 therefore allows dissociation between arousal-driven changes in pitch and stress-related adjustments in phonatory control, clarifying whether stress manifests primarily as vocal degradation or as compensatory tightening of the phonatory system.

Although derived from speech signals, both F0 and NNE index peripheral physiological processes involved in phonation and autonomic regulation, enabling their use as objective, physiology-linked indicators that are largely independent of self-report.

Theoretical rationale. The selection of F0 and NNE was guided by converging neurophysiological and psychophysiological models linking stress to alterations in muscular activation and motor control during speech production (Scherer, 1986; Titze, 1994; Giddens et al., 2013). Stress-related increases in general muscular tone, including at the laryngeal level, are expected to elevate vocal fold tension and vibratory rate, resulting in higher F0. In contrast, predictions for voice quality measures such as NNE are less straightforward. Moderate increases in laryngeal tension may improve glottal adduction and reduce aperiodic energy, whereas excessive or asymmetric tension may stiffen the vocal fold mucosa, disrupt the regularity of the mucosal wave, and increase airflow turbulence at the glottis, leading to higher noise levels (Titze, 1994). This mechanistic ambiguity is consistent with the smaller and less consistent stress effects typically reported for voice quality parameters relative to F0 (Giddens et al., 2013; Van Puyvelde et al., 2018).

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 = 119 participants with complete voice data and valid EMA responses.

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.

Selection of Acoustic Parameters. We focused on two acoustic parameters (F0 and NNE) selected for their theoretical relevance to stress-related vocal changes and their complementary information about distinct physiological mechanisms [REF].

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

Statistical analyses were conducted in a Bayesian framework using Stan. Voice outcomes (mean F0 and NNE) were modeled using hierarchical linear regression models

with repeated observations nested within participants. Experimental effects were parameterized using two orthogonal contrasts representing stress reactivity (stress vs. baseline) and recovery (post-stress vs. stress).

Individual differences in personality were modeled as latent traits derived from repeated EMA assessments of the five PID-5 domains. A multivariate measurement model was used to estimate subject-level latent trait scores while accounting for measurement error in the observed EMA indicators. These latent traits were then included in the outcome model both as main effects and as moderators of stress and recovery effects via cross-level interactions.

All models included subject-specific random intercepts and random slopes for the stress and recovery contrasts. Weakly informative priors were used for all fixed effects, variance components, and correlations. Posterior inference was based on full posterior distributions, with model adequacy evaluated using posterior predictive checks and approximate leave-one-out cross-validation.

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.

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

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. For sustained-vowel outcomes (F0 and NNE), values represent session-level averages computed across the three sustained vowels (/a/, /i/, /u/) and their repetitions.

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 β_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 ($P > 0.99$ that the effects differed from zero for both F0 and NNE) underscore the robustness of these vocal signatures of stress. In contrast, 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 showed meaningful Bayesian evidence. Detachment and Disinhibition exhibited no credible stress-phase moderation (PD = 0.59 and 0.64, respectively). During the recovery phase, Detachment showed a suggestive tendency toward reduced F0 elevation ($\gamma_2 = -2.02$ Hz, PD = 0.88), whereas Disinhibition showed no meaningful effect (PD = 0.67).

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 (γ_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).

Summary. 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 (estimates from repeated EMA assessments) 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

Across an ecologically valid academic stressor, we observed reliable stress-related changes in two complementary vocal parameters, fundamental frequency (F0) and normalized noise energy (NNE), and found *selective, domain- and phase-specific moderation* by PID-5 trait domains. In addition, repeated brief EMA-based trait assessment achieved *comparable predictive accuracy* to a complete baseline PID-5 questionnaire while yielding *greater precision* in key moderation estimates. Together, these findings support transactional accounts of personality pathology by showing that maladaptive trait domains are expressed not only in average tendencies, but also in *how individuals' psychophysiological responses change across situational demands and recovery periods.*

Vocal acoustics as objective indicators of situational stress. Exam-related stress produced dissociable changes in vocal production. Consistent with prior psychophysiological work, F0 increased during the pre-exam assessment, indicating heightened laryngeal muscle tension and autonomic arousal (Giddens et al., 2013; Scherer, 2003). Importantly, this effect emerged in a real-world evaluative context rather than an experimentally induced laboratory stressor, supporting the ecological validity of vocal pitch as an objective marker of stress reactivity in everyday life.

In contrast to a simple "vocal degradation" account, NNE decreased by 0.79 dB during stress, indicating a more periodic, less noisy signal. This pattern suggests that evaluative stress can involve not only arousal-related activation (indexed by F0) but also *compensatory phonatory control* (indexed by reduced glottal noise). Under performance-oriented conditions, individuals may unconsciously adopt more effortful or pressed phonation that increases periodicity, even while physiological arousal rises. This arousal–control dissociation implies that stress responses in the voice are multidimensional

rather than unitary, and that different personality domains may preferentially shape different components of the response.

Notably, average *recovery effects* were weak for both parameters over the short post-exam window. Because exam grades were communicated to participants immediately upon completion (via automated Moodle scoring), the persistence of stress-related vocal changes at T3 cannot be attributed to lingering uncertainty about the outcome. Rather, the absence of clear mean-level recovery likely reflects the slow decay of physiological activation following stress, individual differences in recovery speed, or both. These considerations highlight the value of modeling heterogeneity in recovery trajectories rather than relying solely on mean-level recovery effects. In this sense, recovery may represent a particularly sensitive phase for revealing individual differences that are obscured at the level of average responses.

Domain- and phase-specific moderation of vocal stress responses. Given the absence of prior research on personality moderation of vocal stress markers, we adopted an exploratory approach for most PID-5 domains, specifying a directional hypothesis only for Negative Affectivity. The resulting pattern provides evidence that personality pathology domains do not exert broad, uniform effects; instead, they appear to moderate *specific acoustic parameters at specific phases* of the stress-response cycle.

Negative Affectivity: amplified acute arousal reactivity. As predicted, Negative Affectivity reliably amplified stress-induced increases in F0. Individuals higher in Negative Affectivity showed a stronger pitch elevation during anticipatory stress, consistent with models linking this domain to heightened threat vigilance, emotional lability, and sympathetic activation (Lahey, 2009; Ormel et al., 2013). The specificity of the effect to *F0 during the stress phase*, rather than to NNE or to recovery, supports the interpretation that Negative Affectivity primarily modulates arousal-driven pathways (i.e., the autonomic component of the response), rather than broadly destabilizing vocal output.

Antagonism: evidence for heterogeneous post-stressor trajectories. Antagonism showed its strongest association with the *recovery contrast* for F0, suggesting that individuals higher on antagonistic traits may exhibit different post-stressor trajectories of arousal. Because the average recovery effect in the sample was near zero, this pattern is best interpreted not as a robust population-level "impaired recovery," but as *preliminary evidence of heterogeneity* in the degree to which arousal persists once the acute stressor has ended. One plausible mechanism may involve prolonged engagement with the evaluative episode (e.g., continued anger focus rumination, irritability, or interpersonal conflict in the aftermath), which could maintain physiological activation beyond the stressor offset. However, resolving whether this reflects affective inertia, continued contextual activation, or other processes will require denser post-stressor sampling and measures of post-exam cognitions and contexts.

Psychoticism: selective modulation of voice quality during recovery. Psychoticism showed minimal moderation of F0, aligning with mixed prior evidence on physiological stress reactivity in schizotypy-related traits. However, Psychoticism uniquely moderated *NNE during recovery*, with higher Psychoticism associated with increased glottal noise (less negative NNE) post-exam. This specificity to a voice quality parameter, and to the recovery phase, suggests that Psychoticism may be linked less to the magnitude of acute arousal and more to processes involved in stabilizing phonatory control after stress exposure. One interpretation is that cognitive–perceptual dysregulation may compromise adaptive recalibration of motor control once stress cues recede, yielding less stable phonation. Alternative explanations are also plausible, including persistent autonomic dysregulation or context-dependent changes in speech effort. Discriminating among these accounts will require concurrent measurement of autonomic markers (e.g., heart rate variability) and, ideally, more continuous sampling across the recovery period.

Detachment and Disinhibition: limited evidence for moderation. Detachment showed weak evidence for *improved* rather than impaired F0 recovery (PD =

0.88, negative coefficient). This pattern is noteworthy because it runs counter to predictions derived from the emotional inertia literature, which links restricted affectivity and anhedonia to prolonged negative states and impaired affect repair (Koval et al., 2015; Kuppens et al., 2010). One possible interpretation is that the emotional blunting characteristic of Detachment may attenuate not only positive emotional experiences but also sustained physiological activation following stress: if high-Detachment individuals are less engaged with the evaluative implications of the stressor, they may show faster normalization simply because there is less activation to resolve. Given the modest strength of this effect, it should be treated cautiously and may represent sampling variability, but it raises the possibility that restricted affectivity could function as a context-specific protective factor in acute stress—even if it reflects broader affective impairment in other domains.

Disinhibition showed no meaningful moderation of either acoustic parameter (all PD < 0.75). This null finding is theoretically coherent rather than merely uninformative. Unlike Negative Affectivity, which reflects arousal sensitivity and emotional reactivity, Disinhibition is defined by impulsivity, irresponsibility, and poor behavioral control—characteristics pertaining to volitional action regulation rather than automatic physiological processes (Krueger et al., 2012). Meta-analytic evidence indicates that conscientiousness-related traits show weak associations with physiological stress responses despite moderating subjective stress perception and coping behaviors (Luo et al., 2023). The present findings extend this dissociation to vocal acoustics: because F0 elevation reflects automatic autonomic mechanisms (laryngeal tension from sympathetic activation) rather than behavioral control processes, Disinhibition—which governs volitional rather than automatic responses—would not be expected to moderate this pathway. More broadly, these results reinforce the conclusion that moderation effects are not ubiquitous across PID-5 domains but are *selective and mechanism-specific*, with different domains operating through distinct psychophysiological pathways (Calà et al., 2025).

Implications for transactional models of personality pathology. The dissociations observed across trait domains (Negative Affectivity, Antagonism, Psychoticism), acoustic parameters (F0 vs. NNE), and temporal phases (stress reactivity vs. recovery) are consistent with transactional models in which personality pathology shapes *how* individuals respond to contextual demands rather than simply predicting static levels of functioning (Bolger & Zuckerman, 1995). In this view, different domains may map onto distinct stress-process mechanisms: Negative Affectivity primarily amplifies arousal reactivity; Antagonism may be linked to persistence of activation following stress; and Psychoticism may be associated with post-stressor instability in control-related vocal processes. Importantly, these interpretations remain provisional and should be tested in designs that incorporate richer characterization of situational features and more continuous measurement during recovery.

At the same time, the present findings suggest that vocal acoustics can contribute to a more differentiated phenotyping of stress responses. Relying solely on pitch-based markers may primarily detect internalizing-related vulnerability (Negative Affectivity), whereas incorporating voice-quality indices may reveal additional processes relevant to cognitive–perceptual dysregulation (Psychoticism) or other domains. Future ambulatory work that models multiple vocal dimensions across multiple stressor types may help build more precise, trait-informed accounts of vulnerability and adaptation.

Methodological contribution: intensive ambulatory trait assessment. A central methodological contribution of this project is to show that brief trait measurements assessed repeatedly via EMA can perform comparably to comprehensive baseline questionnaires for predicting vocal stress trajectories. Across models, out-of-sample predictive performance was similar for EMA-based and baseline-based PID-5 measures, indicating a substantial overlap in the trait variance they capture. However, EMA-based measurement yielded *more precise moderation estimates*, consistent with the idea that repeated sampling can reduce measurement error and stabilize person-level trait estimates.

This distinction is important: even when overall predictive accuracy is equivalent, improved inferential precision can strengthen theory tests of trait-by-context interactions.

These results support the feasibility of combining low-burden repeated trait assessment with objective acoustic markers in longitudinal designs, particularly when comprehensive questionnaires are impractical. At the same time, the present design relied on brief standardized voice recordings rather than continuous passive sensing. Extending this approach to more frequent recordings would allow stronger tests of within-person coupling and more detailed modeling of recovery dynamics.

Limitations and future directions. Several limitations qualify the conclusions. First, the sample consisted of female university students, and we excluded participants with psychiatric disorders requiring treatment. Although PID-5 traits are dimensionally meaningful in community samples, generalization to clinical populations and to males requires replication. Second, although the examination period provides a naturalistic stressor with real stakes, it is not equivalent to interpersonal stressors that are central to many forms of personality pathology. Different stressor classes may yield different vocal signatures and trait moderation patterns. Third, vocal data were collected at only three time points, which constrains the temporal resolution with which stress-related dynamics can be characterized. Although this design captures broad phases of baseline, anticipatory stress, and early recovery, it cannot resolve the finer-grained time course through which vocal parameters return to baseline following stressor offset. As a result, apparent persistence of stress-related effects may reflect delayed physiological recovery, continued contextual activation, or both. Future studies employing denser sampling, particularly in the hours immediately following stress exposure, will be necessary to disentangle these processes and to model recovery trajectories with greater precision. Fourth, vocal markers are biologically plausible indices of stress, but they are not process-pure. Changes in F0 and NNE may reflect multiple mechanisms, including autonomic activation, cognitive load, strategic self-presentation, and speech effort. Future studies should triangulate voice

measures with concurrent physiological indicators (e.g., heart rate variability, electrodermal activity) and with contextual measures (perceived stress, task difficulty, preparedness) to clarify underlying pathways.

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.

American Psychiatric Association. (2014). *Online assessment measures: The personality inventory for DSM-5 (adult).* https://www.psychiatry.org/getmedia/594673a6-1b9b-4298-8b52-c4c652c4a4e2/APA-DSM5TR-ThePersonalityInventoryForDSM5FullVersionAdult.pdf.

Bibbey, A., Carroll, D., Roseboom, T. J., Phillips, A. C., & Rooij, S. R. de. (2013). Personality and physiological reactions to acute psychological stress. *International Journal of Psychophysiology*, *90*(1), 28–36.

Bottesi, G., Caudek, C., Colpizzi, I., Iannattone, S., Palmieri, G., & Sica, C. (2024). Advancing understanding of the relation between criterion a of the alternative model for personality disorders and hierarchical taxonomy of psychopathology: Insights from an external validity analysis. *Personality Disorders: Theory, Research, and Treatment.*

Calà, F., Colpizzi, I., Sica, C., Caudek, C., Lanatà, A., & Frassineti, L. (2025). A preliminary analysis on longitudinal effects of exam stress and personality traits over acoustic properties. *Models and Analysis of Vocal Emissions for Biomedical Applications: 15th International Workshop*, 145–148.

Colpizzi, I., Trull, T. J., Sica, C., Haney, A. M., & Caudek, C. (2025). State self-compassion dynamics: Partial evidence for the bipolar continuum hypothesis. *Mindfulness*, 1–16.

Dietrich, M., & Abbott, K. V. (2012). Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *Journal of Speech, Language, and Hearing Research*, *55*(3), 973–987. https://doi.org/10.1044/1092-4388(2011/10-0344)

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice*, *19*(4), 544–554.

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.

Luo, J., Zhang, B., Cao, M., & Roberts, B. W. (2023). The stressful personality: A meta-analytical review of the relation between personality and stress. *Personality and Social Psychology Review*, *27*(2), 128–194.

Manfredi, C., Lebacq, J., Cantarella, G., Schoentgen, J., Orlandi, S., Bandini, A., & DeJonckere, P. H. (2017). Smartphones offer new opportunities in clinical voice research. *Journal of Voice*, *31*(1), 111–e1.

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

Morelli, M. S., & Manfredi, S. O. C. (2019). BioVoice: A multipurpose tool for voice analysis. *Proceedings of the 11th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA 2019*, 261–264. Firenze University Press Firenze, Italy.

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

Sica, C., Caudek, C., Colpizzi, I., Bottesi, G., Iannattone, S., & Patrick, C. J. (2024). Comparing the DSM-5 dimensional trait and triarchic model conceptions of psychopathy: An external validity analysis. *Journal of Personality Disorders*, *38*(4), 368–400.

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

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), 56–63. https://doi.org/10.1037/abn0000473

Van Puyvelde, M., Neyt, X., McGlone, F., & Pattyn, N. (2018). Voice stress analysis: A new framework for voice and effort in human performance. *Frontiers in Psychology*, *9*, 1994.

Table 1

Principal Acoustic Features of Speech: Definitions and Psychological Significance

Feature	Definition	Physiological basis	Psychological significance
Fundamental frequency (F0)	Rate of vocal fold vibration (Hz); acoustic correlate of perceived pitch	Cricothyroid and vocalis muscle tension regulates vocal fold stiffness and vibratory rate	Most robust acoustic marker of stress: F0 increases reliably under acute stress, cognitive load, and evaluative threat via sympathetic activation
Normalized Noise Energy (NNE)	Ratio of inharmonic to harmonic spectral energy (dB); index of glottal noise	Degree of glottal closure completeness during phonation	Reflects phonatory control and vocal effort; stress may increase noise (incomplete closure under arousal) or decrease it (compensatory hyperadduction and pressed phonation)
Jitter[a]	Cycle-to-cycle variation in F0 period (%); measure of frequency perturbation	Irregularity of successive vocal fold vibratory cycles	Proposed as an index of vocal instability and reduced neuromuscular control; elevated in some emotional and clinical states

Note. F0 and NNE were retained as primary outcome variables in the present study. [a]Jitter was extracted but not included in the main analyses for the following reasons: (a) the literature on jitter and psychological stress is inconsistent, with effect sizes that are small, heterogeneous in direction, and often non-significant across studies (Giddens et al., 2013); (b) jitter is highly sensitive to recording conditions, signal-to-noise ratio, and extraction algorithm parameters, resulting in substantial measurement noise in field settings (Titze, 1994); and (c) unlike F0 and NNE, which index theoretically distinct mechanisms (autonomic arousal vs. phonatory control), jitter does not map onto a clearly separable psychophysiological pathway, limiting its interpretive value in the context of personality moderation hypotheses.

Table 2

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 3

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 4

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 5

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.