# 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

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline. Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines. One sentence clearly stating the **general problem** being addressed by this particular study. One sentence summarizing the main result (with the words "**here we show**" or their equivalent). Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

*Keywords:* keywords

Word count: X

The title

## Descriptive statistics

Descriptive statistics are reported for the **final analysis sample** (female participants only; $N = 119$), using the same datasets entered in the Stan models for F0 and NNE. Participants contributed on average 27 EMA assessments ($SD = 4.20$, range = 12–31).

**Design and coverage.** All 119 participants contributed usable voice observations in each assessment period (Baseline, Pre-exam, Post-exam), consistent with the balanced, three-phase design.

**Voice outcomes (F0 and NNE) by period.** We report voice descriptives at the **observation level** (all available voice observations per period; Table S2a) and at the **between-person level** using **person means** within each period (Table S2b). In this dataset, each participant contributed one observation per period, therefore Tables S2a and S2b coincide.

Across periods, mean F0 increased from Baseline to Pre-exam and decreased from Pre-exam to Post-exam, consistent with an anticipatory stress-related elevation in pitch followed by partial recovery. NNE showed a shift (more negative values) from Baseline to Pre-exam and a partial return toward Baseline in the Post-exam period.

To summarize **within-person variability** across the three periods, Table S2c reports the distribution of participant-specific standard deviations computed across Baseline, Pre-exam, and Post-exam.

**EMA personality domain descriptives (PID-5).** EMA-based PID-5 domain descriptives are reported separately for (a) **between-person** variability, computed from each participant's mean across EMA occasions (Table S3a), and (b) **within-person** variability, summarized as the distribution of each participant's within-person standard deviation across EMA occasions (Table S3b).

Between-person descriptives indicate substantial inter-individual variability in EMA trait levels, while within-person descriptives confirm meaningful intra-individual fluctuation across repeated assessments—consistent with the multilevel measurement model used to estimate latent trait scores and propagate measurement uncertainty into moderation effects.

# Reliability of Personality Measures

## Overview

Reliability was evaluated for both the full baseline PID-5 questionnaire (220 items) and the brief EMA-based PID-5 assessment (15 items, 3 per domain). For the EMA measures, which involve repeated assessments nested within persons, we employed multilevel reliability estimation following Lai (2021), which distinguishes between-person reliability (consistency of person-level means) from within-person reliability (consistency of occasion-level deviations).

## Baseline PID-5 Questionnaire

The full PID-5 was administered at baseline (T1) using the standard 220-item version. Reliability was computed using Cronbach's alpha and McDonald's omega for each domain and for the total score. Results are reported for both the full sample and a cleaned sample excluding participants who failed attention check items (catch trials embedded at positions 68 and 161 in the questionnaire).

**Results.** All domains showed acceptable to excellent internal consistency. Psychoticism exhibited the highest reliability ($\alpha = .87$, $\omega = .90$), consistent with its larger item pool (33 items). Antagonism showed the lowest reliability ($\alpha = .67$, $\omega = .82$), though McDonald's omega—which accounts for item heterogeneity—indicated adequate composite reliability. The total PID-5 score demonstrated excellent reliability ($\alpha = .96$, $\omega = .98$). Excluding participants who failed attention checks produced slightly lower but comparable estimates, indicating that careless responding had minimal impact on scale properties in this sample.

## EMA-Based Brief PID-5

The brief PID-5 administered via EMA comprised 15 items (3 per domain), selected based on factor loadings and domain representativeness from prior validation work. Because these items were assessed repeatedly across approximately 20 occasions per participant, standard single-level reliability estimates are inappropriate. Instead, we employed the multilevel composite reliability framework described by Lai (2021), which partitions variance into within-person and between-person components and computes separate reliability indices for each level.

**Multilevel Reliability Framework.** Following Lai (2021), we estimated three reliability indices:

- $\alpha_{2L}$ **(Two-level alpha)**: Overall reliability of observed scores, pooling within- and between-person variance. Relevant when scores are used without distinguishing levels.
- $\alpha_B$ **(Between-person alpha)**: Reliability of person-level means (aggregated across occasions). This is the relevant index when EMA scores are averaged to create a single trait estimate per person, as in our moderation analyses.
- $\alpha_W$ **(Within-person alpha)**: Reliability of occasion-level deviations from each person's mean. Relevant for detecting momentary fluctuations in personality states.

The same decomposition was applied using McDonald's omega ($\omega_{2L}$, $\omega_B$, $\omega_W$), which relaxes the assumption of tau-equivalence.

**Results.** The multilevel reliability analysis revealed a clear pattern:

1. **Between-person reliability was high** ($\alpha_B = .87$, $\omega_B = .85$). This indicates that person-level trait estimates derived from aggregating across EMA occasions are measured with good precision. This is the critical index for our moderation analyses, which use person-level latent trait scores as predictors.

2. **Within-person reliability was adequate** ($\alpha_W = .73$, $\omega_W = .72$). This suggests that the brief scale can detect meaningful occasion-to-occasion fluctuations in personality states, though with more measurement error than at the between-person level. This is expected given the brevity of the scale (15 items total).

3. **Two-level reliability was good** ($\alpha_{2L} = .83$, $\omega_{2L} = .81$), reflecting the overall quality of observed scores when levels are pooled.

**Interpretation for the Present Study.** The high between-person reliability ($\alpha_B = .87$) supports the validity of using EMA-derived personality scores as person-level moderators of vocal stress responses. Because our moderation model estimates latent trait scores from the repeated EMA observations (see Statistical Models section), measurement error is explicitly modeled rather than ignored. The multilevel measurement model in our Stan implementation can be viewed as formalizing the reliability structure documented here: the latent trait $\theta_{id}$ represents the "true" person-level standing on domain $d$, while the occasion-specific observations $X_{nd}$ are treated as noisy indicators with residual variance $\sigma_d^{\text{ema}}$ capturing within-person fluctuation and measurement error.

The adequate within-person reliability ($\alpha_W = .73$) also suggests that the brief EMA measure could support analyses of state-level personality-voice covariation, though such analyses would require denser sampling than the present design to achieve adequate statistical power.

# Statistical Models

## Main Effects of Exam-Related Stress on Vocal Parameters

**Research Question.** Before examining personality moderation, we first establish whether exam-related stress produces reliable changes in vocal acoustics. This model addresses the foundational question: Do vocal parameters (F0, NNE) change systematically across the three assessment phases—baseline, pre-exam (anticipatory stress), and post-exam (recovery)?

**Model Specification.** The model is a standard Bayesian hierarchical (multilevel) linear regression with repeated measures nested within participants. Stress effects are parameterized using two orthogonal contrasts:

- **Stress contrast ($c_1$)**: Compares pre-exam to baseline, capturing anticipatory stress-induced change.
- **Recovery contrast ($c_2$)**: Compares post-exam to pre-exam, capturing post-stressor trajectory.

Let $y_n$ denote the vocal outcome for observation $n$, with participant index $s[n]$. The model specifies:

$$y_n \sim \text{Normal}(\mu_n, \sigma_y).$$

where the linear predictor includes both fixed and random effects:

$$\mu_n = \alpha + u_{0,s} + (\beta_1 + u_{1,s}) \cdot c_{1n} + (\beta_2 + u_{2,s}) \cdot c_{2n}.$$

The parameters are:

- $\alpha$: Grand intercept (population-average baseline vocal level).
- $\beta_1$: Population-average stress effect (pre-exam vs. baseline).
- $\beta_2$: Population-average recovery effect (post-exam vs. pre-exam).
- $u_{0,s}$: Participant-specific random intercept.
- $u_{1,s}$: Participant-specific random slope for stress.
- $u_{2,s}$: Participant-specific random slope for recovery.
- $\sigma_y$: Residual standard deviation.

The random effects capture individual differences in baseline vocal characteristics ($u_0$), stress reactivity ($u_1$), and recovery patterns ($u_2$). A non-centered parameterization is used for computational efficiency:

$$u_{k,s} = \tau_k \cdot z_{k,s}, \quad z_{k,s} \sim \text{Normal}(0, 1).$$

where $\tau_k$ are the random effect standard deviations.

### Prior Specification.

```
data {
  int<lower=1> N_subj;                            // number of subjects
  int<lower=1> N_obs;                             // total observations
  array[N_obs] int<lower=1, upper=N_subj> subj_id; // subject index per observation
  vector[N_obs] y;                                // F0 mean (Hz)
  vector[N_obs] c1;                               // stress contrast: PRE vs BASELINE
  vector[N_obs] c2;                               // recovery contrast: POST vs PRE
}

parameters {
  // Fixed effects
  real alpha;                   // grand intercept (baseline F0)
  real b1;                      // stress main effect (c1)
  real b2;                      // recovery main effect (c2)

  // Random effects (non-centered parameterization)
  vector[N_subj] z_u0;          // random intercepts (standardized)
  vector[N_subj] z_u1;          // random slopes for c1 (standardized)
  vector[N_subj] z_u2;          // random slopes for c2 (standardized)
  vector<lower=0>[3] tau;       // SDs: tau[1]=intercept, tau[2]=c1, tau[3]=c2

  real<lower=0> sigma_y;        // residual SD
}

transformed parameters {
  // Non-centered random effects
  vector[N_subj] u0 = tau[1] * z_u0;
  vector[N_subj] u1 = tau[2] * z_u1;
  vector[N_subj] u2 = tau[3] * z_u2;
}

model {
  // Priors
  alpha ~ normal(220, 30);
  b1 ~ normal(0, 10);
  b2 ~ normal(0, 10);

  tau ~ exponential(0.5);
  sigma_y ~ exponential(0.1);

  z_u0 ~ std_normal();
  z_u1 ~ std_normal();
  z_u2 ~ std_normal();

  // Likelihood
  for (n in 1:N_obs) {
    int s = subj_id[n];
    real mu = alpha + u0[s]
            + (b1 + u1[s]) * c1[n]
            + (b2 + u2[s]) * c2[n];
    y[n] ~ normal(mu, sigma_y);
  }
}
```

```
generated quantities {
  // Posterior predictive replicates for model checking
  vector[N_obs] y_rep;

  // Log-likelihood for model comparison (LOO, WAIC)
  vector[N_obs] log_lik;

  for (n in 1:N_obs) {
    int s = subj_id[n];
    real mu = alpha + u0[s]
            + (b1 + u1[s]) * c1[n]
            + (b2 + u2[s]) * c2[n];

    y_rep[n] = normal_rng(mu, sigma_y);
    log_lik[n] = normal_lpdf(y[n] | mu, sigma_y);
  }
}
```

**Stan Implementation.**

**Interpretation.** The key parameters of interest are $\beta_1$ (stress effect) and $\beta_2$ (recovery effect):

- A positive $\beta_1$ indicates that F0 increases from baseline to pre-exam, consistent with heightened autonomic arousal elevating vocal pitch.
- A negative $\beta_2$ would indicate that F0 decreases from pre-exam to post-exam, suggesting recovery toward baseline levels.

The random effect standard deviations ($\tau_1$, $\tau_2$, $\tau_3$) quantify the degree of individual differences in baseline levels, stress reactivity, and recovery patterns. Large values of $\tau_2$ or $\tau_3$ would indicate substantial heterogeneity in how participants respond to stress—heterogeneity that might be explained by personality traits, motivating the moderation analyses presented in the next section.

The `generated quantities` block produces posterior predictive samples for model checking and pointwise log-likelihoods for model comparison via LOO-CV.

---

## Personality Moderation of Vocal Stress Responses

**Research Question.** The central question addressed by this model is whether the effect of exam-related stress on vocal acoustics (F0, NNE) is moderated by individual differences in personality pathology. Specifically, we ask: Do the five PID-5 domains—Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism—differentially amplify or attenuate the stress-induced changes in vocal parameters during (a) anticipatory stress and (b) post-stressor recovery?

**The Challenge: Personality Traits from Intensive Longitudinal Data.** A key methodological challenge in this study concerns how personality traits are represented in the moderation analysis. Each participant completed approximately 20 EMA assessments over 2.5 months, providing repeated measures of each PID-5 domain. A naive approach would aggregate these observations into a single person-level mean for each domain and use these means as predictors in a standard multilevel regression. However, this approach discards valuable information and fails to account for measurement error: the observed person-means are noisy estimates of the true latent traits, and treating them as known quantities underestimates uncertainty in the moderation effects.

Our model addresses this problem through a *joint measurement-and-outcome model* that simultaneously estimates latent personality traits from the EMA data and their moderating influence on vocal stress responses. This integrated approach has three key advantages:

1. **Measurement error correction**: Rather than using observed means as fixed predictors, the model estimates each participant's true latent trait score ($\theta_{id}$) as a parameter, with appropriate uncertainty. This uncertainty propagates into the moderation estimates, yielding appropriately calibrated credible intervals.

2. **Borrowing strength across observations**: The repeated EMA measurements inform the latent trait estimates through a measurement model, allowing the model to distinguish stable trait variance from occasion-specific fluctuations.

3. **Coherent uncertainty quantification**: Because the latent traits and their effects are estimated jointly, the posterior distributions for moderation parameters ($\gamma_1$, $\gamma_2$) fully reflect uncertainty about both the traits themselves and their influence on vocal outcomes.

**Model Specification.** The model comprises two interconnected components: a *measurement model* for the EMA-based personality assessments and an *outcome model* for the vocal parameters.

***Measurement Model (EMA).*** Let $X_{nd}$ denote the observed score for participant $i$ on domain $d$ at EMA occasion $n$. The measurement model specifies:

$$X_{nd} \sim \text{Normal}(\theta_{i[n],d}, \sigma_d^{\text{ema}}),$$

where $\theta_{id}$ is the latent true trait score for participant $i$ on domain $d$, and $\sigma_d^{\text{ema}}$ captures occasion-to-occasion variability (including both state fluctuations and measurement error). The latent traits are given standard normal priors:

$$\theta_{id} \sim \text{Normal}(0, 1).$$

This formulation treats each participant's approximately 20 EMA observations as repeated noisy indicators of a stable underlying trait, with the model learning both the trait estimates and the amount of occasion-level variability for each domain.

***Outcome Model (Vocal Parameters).*** Let $y_j$ denote the vocal outcome (F0 or NNE) for observation $j$, with participant index $i[j]$. Stress effects are parameterized using two orthogonal contrasts:

- $c_1$: Stress contrast (pre-exam vs. baseline);
- $c_2$: Recovery contrast (post-exam vs. pre-exam).

The outcome model specifies:

$$y_j \sim \text{Normal}(\mu_j, \sigma_y),$$

where the linear predictor $\mu_j$ includes fixed effects, random effects, and the crucial trait × contrast interactions:

$$\mu_j = \alpha + u_{i,1} + (\beta_1 + u_{i,2}) \cdot c_{1j} + (\beta_2 + u_{i,3}) \cdot c_{2j} + \sum_{d=1}^{5} \left[a_d \cdot \theta_{id} + \gamma_{1d} \cdot c_{1j} \cdot \theta_{id} + \gamma_{2d} \cdot c_{2j} \cdot \theta_{id}\right].$$

The parameters are:

- $\alpha$: Grand mean (baseline vocal parameter);
- $\beta_1$, $\beta_2$: Population-average stress and recovery effects;
- $u_{i,1}$, $u_{i,2}$, $u_{i,3}$: Participant-specific random intercept and slopes;
- $a_d$: Main effect of trait $d$ on baseline vocal level;
- $\gamma_{1d}$: **Stress moderation**—how trait $d$ amplifies or attenuates the stress effect;
- $\gamma_{2d}$: **Recovery moderation**—how trait $d$ shapes post-stressor trajectory.

The moderation parameters $\gamma_{1d}$ and $\gamma_{2d}$ are the quantities of primary theoretical interest. A positive $\gamma_{1d}$ indicates that higher levels of trait $d$ are associated with larger stress-induced changes in the vocal parameter.

***Random Effects Structure.*** Participant-level random effects are specified using a non-centered parameterization for computational efficiency:

$$u_{i,k} = z_{i,k} \cdot \tau_k, \quad z_{i,k} \sim \text{Normal}(0, 1),$$

where $\tau_k$ are the random effect standard deviations. This structure allows for individual differences in baseline levels ($u_{i,1}$), stress reactivity ($u_{i,2}$), and recovery ($u_{i,3}$) beyond what is explained by the PID-5 traits.

**Prior Specification.** Priors were chosen to be weakly informative, incorporating domain knowledge about plausible parameter ranges while allowing the data to dominate inference:

The priors on the moderation parameters ($\gamma_{1d}$, $\gamma_{2d}$) provide modest regularization, shrinking estimates toward zero in the absence of strong evidence. This helps guard against overfitting given the 10 moderation parameters (5 domains × 2 contrasts) being estimated.

**Stan Implementation.** The model was implemented in Stan. The complete code is provided below.

```
data {
  int<lower=1> N_subj;

  // Voice outcome
  int<lower=1> N_voice;
  array[N_voice] int<lower=1, upper=N_subj> subj_voice;
  vector[N_voice] y;
  vector[N_voice] c1;  // stress contrast
  vector[N_voice] c2;  // recovery contrast

  // EMA measurement model
  int<lower=1> N_ema;
  array[N_ema] int<lower=1, upper=N_subj> subj_ema;
  int<lower=1> D;              // 5 domains
  matrix[N_ema, D] X;          // standardized EMA domain scores
}

parameters {
  // Latent traits (true person means): theta[i,d]
  matrix[N_subj, D] theta;
  vector<lower=0>[D] sigma_ema;

  // Fixed effects for voice
  real alpha;          // grand intercept
  real b1;             // stress main effect
  real b2;             // recovery main effect

  // Main effects of traits on baseline voice (optional)
  vector[D] a_trait;

  // Moderation: trait × stress and trait × recovery
  vector[D] g1;        // stress moderation (c1 * theta)
  vector[D] g2;        // recovery moderation (c2 * theta)

  // Random effects (no correlations): intercept + slopes
  vector<lower=0>[3] tau;      // SDs for random intercept, stress slope, recovery slop
  matrix[N_subj, 3] z_u;       // standard normals
  real<lower=0> sigma_y;       // residual SD
}

transformed parameters {
  matrix[N_subj, 3] u;
  u = z_u;
  for (i in 1:N_subj) {
    u[i,1] = u[i,1] * tau[1];
    u[i,2] = u[i,2] * tau[2];
    u[i,3] = u[i,3] * tau[3];
  }
}

model {
  // -------------------------
  // Priors
  // -------------------------
  to_vector(theta) ~ normal(0, 1);
  sigma_ema ~ exponential(1);

  alpha ~ normal(220, 30);
  b1 ~ normal(0, 10);
  b2 ~ normal(0, 10);

  a_trait ~ normal(0, 5);
```

```
// moderation: shrinkage (second-order)
g1 ~ normal(0, 3);
g2 ~ normal(0, 3);

tau ~ exponential(0.5);
to_vector(z_u) ~ normal(0, 1);

sigma_y ~ exponential(0.1);

// --------------------------
// Measurement model (EMA)
// --------------------------
for (n in 1:N_ema) {
  for (d in 1:D) {
    X[n,d] ~ normal(theta[subj_ema[n], d], sigma_ema[d]);
  }
}

// --------------------------
// Voice outcome model
// --------------------------
for (j in 1:N_voice) {
  int i = subj_voice[j];

  real mu = alpha
    + u[i,1]
```

```
      + (b1 + u[i,2]) * c1[j]
      + (b2 + u[i,3]) * c2[j];

    for (d in 1:D) {
      mu += a_trait[d] * theta[i,d]
          + g1[d] * c1[j] * theta[i,d]
          + g2[d] * c2[j] * theta[i,d];
    }

    y[j] ~ normal(mu, sigma_y);
  }
}

generated quantities {
  vector[N_voice] y_rep;
  for (j in 1:N_voice) {
    int i = subj_voice[j];
    real mu = alpha + u[i,1] + (b1 + u[i,2]) * c1[j] + (b2 + u[i,3]) * c2[j];
    for (d in 1:D) {
      mu += a_trait[d] * theta[i,d]
          + g1[d] * c1[j] * theta[i,d]
          + g2[d] * c2[j] * theta[i,d];
    }
    y_rep[j] = normal_rng(mu, sigma_y);
  }
}
```

**Interpretation of Key Parameters.** The model yields posterior distributions for 10 moderation parameters of primary interest:

- $\gamma_{1,\mathbf{NegAff}}, \ldots, \gamma_{1,\mathbf{Psych}}$: How each PID-5 domain moderates stress-induced vocal change (stress contrast × trait).
- $\gamma_{2,\mathbf{NegAff}}, \ldots, \gamma_{2,\mathbf{Psych}}$: How each domain moderates recovery-phase vocal change (recovery contrast × trait).

These parameters are expressed in the original units of the vocal outcome (Hz for F0, dB for NNE) per standard deviation of the latent trait. For example, $\gamma_{1,\text{NegAff}} = 3.0$ would indicate that a one-SD increase in latent Negative Affectivity is associated with an additional 3 Hz increase in F0 during the stress phase, beyond the population-average stress effect.

**Posterior Predictive Checks.** The `generated quantities` block produces posterior predictive samples ($y_{\text{rep}}$) for model checking. These were used to verify that the model adequately captured the distributional properties of the observed vocal data, including means, variances, and the pattern of individual differences across assessment phases.

# Posterior Predictive Assurance of Moderation Effects

## Rationale

Posterior predictive simulations were conducted for the F0 outcome to estimate the probability that the direction of the observed moderation effect would replicate under the same study design. Beyond estimating posterior distributions for the moderation parameters, we conducted a posterior predictive assurance analysis to quantify the probability that the *direction* of the observed moderation effect would replicate under the same study design. Rather than addressing statistical significance or interval exclusion criteria, this analysis focuses on directional replicability: given the fitted model and the uncertainty in its parameters, how likely is it that a new study with the same design would yield a moderation effect in the same direction?

This approach is conceptually distinct from frequentist power analysis. Instead of assuming a fixed but unknown true effect size, posterior predictive assurance integrates over the full posterior distribution of the model parameters, thereby reflecting uncertainty about both the magnitude of the effect and the data-generating process.

## Procedure

Posterior predictive simulations were conducted using draws from the joint posterior distribution of the fitted moderation model. For each simulation, a new dataset was generated under the same design as the original study, including the same number of participants, the same stress and recovery contrasts, and a comparable distribution of EMA observations per participant. Latent personality traits for the new participants were drawn from their population prior distributions, consistent with the generative assumptions of the model.

For each simulated dataset, the moderation effect of interest—specifically, the interaction between Negative Affectivity and the stress contrast—was re-estimated using a fast proxy model. Replication success was defined using a minimal and directionally focused criterion: the estimated moderation effect was required to be positive ($> 0$). This criterion captures whether the effect would replicate in direction, without imposing additional thresholds related to statistical significance or effect size magnitude. This process was repeated 1,000 times, yielding an empirical estimate of the posterior predictive probability of directional replication (assurance).

## Results

Across posterior predictive replications, the moderation effect was positive in 92.1% of simulated datasets. The estimated posterior predictive probability of replication success was therefore 0.92, with a 95% credible interval of [0.90, 0.94], reflecting Monte Carlo uncertainty in the simulation-based estimate.

## Interpretation

These results indicate a high probability that the direction of the moderation effect would replicate in a new sample drawn under the same design assumptions. In other words, given the fitted model and the uncertainty in its parameters, the interaction between Negative Affectivity and stress is expected to be positive in the large majority of replications.

At the same time, this analysis does not imply precise recovery of the effect magnitude. The posterior distribution of the moderation parameter remains relatively wide, indicating substantial uncertainty about the exact size of the effect. The assurance analysis therefore supports *directional robustness* of the moderation effect, while remaining agnostic about the degree of precision with which its magnitude can be estimated.

Taken together with the posterior estimates reported in the main analysis, these results suggest that the observed moderation effect is unlikely to be a chance reversal in direction, even though its quantitative strength should be interpreted with appropriate caution.

## References

| Parameter | Prior | Rationale |
|---|---|---|
| $\alpha$ | Normal(220, 30) | Centered on typical female F0 (Hz) |
| $\beta_1$, $\beta_2$ | Normal(0, 10) | Weakly informative; allows effects up to ±20 Hz |
| $\tau_k$ | Exponential(0.5) | Weakly informative for random effect SDs |
| $\sigma_y$ | Exponential(0.1) | Allows residual SD in plausible range |

| Parameter | Prior | Rationale |
|---|---|---|
| $\alpha$ | Normal(220, 30) | Centered on typical female F0 (Hz) |
| $\beta_1$, $\beta_2$ | Normal(0, 10) | Allows stress effects up to ±20 Hz |
| $a_d$ | Normal(0, 5) | Modest trait effects on baseline |
| $\gamma_{1d}$, $\gamma_{2d}$ | Normal(0, 3) | Regularization toward zero |
| $\tau_k$ | Exponential(0.5) | Weakly informative for SDs |
| $\sigma_y$ | Exponential(0.1) | Allows residual SD up to ~10 Hz |
| $\sigma_d^{\text{ema}}$ | Exponential(1) | Weakly informative for EMA variability |