Q1 Analytical questionnaire This questionnaire is for analysts in the Many Speech Analyses project.

Please provide us with information about your research workflow and other related details.

We used a combination of Github and OSF as our research workflow. Each member pushed and pulled from the Github (<https://github.com/doomlab/many_speeches>) which automatically synced with our OSF account (<https://osf.io/ycn9r/>). Further, we set up bi-weekly meetings (as needed) and discussed the selection of the dependent variable, the structure of the analyses, and how to prepare the information for sharing the final product.

Q2 Which software did you use for the processing of the recordings and extraction of acoustic measurements? Please indicate all pieces of software that were used, and specify what was accomplished with each. For example: "Filtering of the recordings was done using the Matlab signal processing toolbox, while the extraction of acoustic measurements (f0, duration, formants, CoG) was accomplished with Praat."

The processing of the recordings and the extraction of acoustic measurements were done with Praat [praat6209](https://www.fon.hum.uva.nl/praat/praat6209_mac.dmg), R version 4.1.1, and rPraat v. 1.3.2.1 package.

Q3 Which analytical/statistical technique(s) did you employ to answer the research question? Please, indicate the name of the techniques and describe specific details if necessary. For example, "We run Bayesian mixed-effects models using a Gaussian likelihood as the Page 2 of 6 distribution of the outcome variable". Details about variables and parameters will be asked in later questions. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

We fitted a linear mixed model (estimated using REML and nloptwrap optimizer) using the lme4 package in R.

Q4 How familiar are you with the analytical/statistical technique you employed?

Jason: Very familiar

Erin: Super Duper Familiar

Ryan: Slightly familiar

Q4 To answer the research question, which variable(s) did you choose as the dependent variable(s) (outcome or measured or latent variable), and what was the rationale behind the choice? If the analysis included multiple dependent variables, specify all of them and the rationale behind choosing each.

We used intensity at the segmental level as our DV. Fundamental frequency, duration, and intensity are all commonly cited correlates of stress in languages such as English and German (Baumann et al., 2007; Braun, 2006; Braun & Ladd, 2003; Katz & Selkirk, 2011; Mixdorff et al., 2015; Sityaev & House, 2003; for a review, see Gordon & Roettger, 2017). We chose intensity both because it is a reliable indicator of stress and contrastive focus, and because of the difficulty extracting durational information from the raw recordings. We extracted peak intensity values from each sentence under the hypothesis that the target word would be subject to contrastive focus and likely show increased duration and intensity, as well as elevated F0.

Q5 To answer the research question, which variable(s) did you choose as the independent variable(s) (predictors), and what was the rationale behind the choice? If the analysis included multiple independent variables, specify all of them and the rationale behind choosing each.

The only IV we included was typicality (as a continuous measure). This was based on the question of interest (i.e., Do speakers phonetically modulate utterances to signal atypical word combinations?).

We chose to use the continuous measure of typicality, over the categorical option, because we believed it provided more information in the continuous form to explore information about the nature of typicality across the spectrum. Further, this provided one parameter estimate (*b*) for typicality predicting our DV, rather than categorical comparisons between each level of typicality.

We also included the random intercept of participants (filename below) to control for the repeated measurement of participants which was cross classified with the item target name.

Intensity ~ typ\_mean + (1|target\_name) + (1|filename) + (1|filename:target\_name)

How did you operationalise your dependent and independent variables? For example, if you choose f0 as a variable, specify from which speech domain it was extracted (for example: the phrase, the word, the syllable, the vowel, etc.), whether you extracted a summary statistic over a larger domain or a concrete value at a specific or relative point in time. For categorical variables (like age, gender, etc.), specify the levels/categories, how they are operationalised and the rationale behind the categorisation.

IV: Typicality was treated as a continuous measure, as provided in the dataset (0 to 100 typical features of the object).

DV: We extracted intensity in dB from the segment level. We broke down each segment into (approximately) milliseconds to find the maximum intensity for the segment.

Q7 What transformation (if any) were applied to the variables? Please, be as specific as possible by indicating for each variable that has been transformed the name of the variable as found in the dataset and the formula/operation that has been applied to it. For example: "We log transformed the F0 values, which are found in the column 'log\_f0' of our data frame."

We did not transform any of our variables.

Q8 Were any observations excluded? If so, which criteria did you follow for the exclusion? For example: "We excluded values that were +/- 3 standard deviations from the mean."

We did not exclude any observations from the data.

Q9 What criteria did you use to decide on an answer to the research question? For example, the decision could have been based on p-values for particular predictors and/or interactions, on Bayesian posterior distributions, or confidence/credible intervals, etc…

We used p < .05 as our criteria.

Q10 What is the size of the effect for the research question? Please, specify the magnitude and direction of the effect size, along with the 95% confidence (or credible) interval in the following format: estimate [low interval, high interval].

Zero

< .001 [.000, .999]

Q11 What unit is your effect size in (i.e., Hedges' G, standardized mean difference, etc.)?

Pseudo-R2 < .000

Q12 Often we do not anticipate possible road blocks during our analyses. Did you encounter scenarios in which you had to revise an earlier analytical decision at a later stage, e.g. change a measurement after you have statistically analyzed (parts of) the data? If so, please briefly describe the situation.

We originally discussed using typicality as a categorical variable. This model was analyzed, but then we figured out the typicality mean was included in the data, and we switched to the more descriptive version of typicality. Additionally, we discussed that the team likely wanted one parameter estimate, but the categorical option would create multiple parameter estimates (i.e., 3 *b* values if you compared all versions of typicality).