



Data analysis questionnaire

Start of Block: Default Question Block

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.

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."

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



distribution of the outcome variable". Details about variables and parameters will be asked in later questions.

Q13 How familiar are you with the analytical/statistical technique you employed?

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.



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.

Q6 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.



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."

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."

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...



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].

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



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.

End of Block: Default Question Block
