Voice paper

March 2022

Contents

Supplemental Methods

Statistical analysis

Model I: individual and recording signal

We used the dis-similarity matrices from DTW, SPCC, MFCCCC or SPECAN to get a pairwise acoustic distance between all calls. We only included call pairs from the same year. We used a model similar to the Bayesian social relations model used in psychology. In our case the acoustic distance between any two calls was the response variable and was standardised before analysis. We included a global intercept and off-sets for: same vs different individual, same vs different recording, pair of individuals from which the calls came, pair of recordings from which the calls came and each call. The full model structure was as follows:

$$
\begin{aligned}
\text{accoustic distance} &\sim \text{normal}(\mu,\ \sigma) \\
\mu_n &= \bar{\alpha} + \alpha_{\text{same ind[n]}} + \alpha_{\text{same rec[n]}} + \\
&\quad \alpha_{\text{ind pair[n]}} + \alpha_{\text{rec pair[n]}} + \alpha_{\text{call i[n]}} + \alpha_{\text{call i[n]}} \\
\bar{\alpha} &\sim \text{normal}(0,\ 0.25) \\
\alpha_{\text{same ind}} &\sim \text{normal}(0,\ \sigma_{\text{same ind}}) \\
\alpha_{\text{same rec}} &\sim \text{normal}(0,\ \sigma_{\text{same rec}}) \\
\alpha_{\text{ind pair}} &\sim \text{normal}(0,\ \sigma_{\text{ind pair}}) \\
\alpha_{\text{rec pair}} &\sim \text{normal}(0,\ \sigma_{\text{rec pair}}) \\
\alpha_{\text{call}} &\sim \text{normal}(0,\ \sigma_{\text{call}}) \\
\sigma_{\text{same ind}},\ \sigma_{\text{same rec}},\ \sigma_{\text{ind pair}},\ \sigma_{\text{rec pair}} &\sim \text{exponential}(3) \\
\sigma,\ \sigma_{\text{call}} &\sim \text{exponential}(5)
\end{aligned}
$$

We fitted an un-centred version of the model using the package *cmdstanr* (Gabry and Češnovar 2021) with the No U-turn Sampler in Stan (Gelman, Lee, and Guo 2015) in from R (R Core Team 2021). We ran 2000 iterations on four chains with adapt_delta = 0.99 and max_treedepth = 15. We used the difference between $\alpha_{\text{same ind [same]}}$ and $\alpha_{\text{same ind [different]}}$ as measure of individual signal and the difference between $\alpha_{\text{same rec [same]}}$ and $\alpha_{\text{same rec [different]}}$ as measure of recording signal. The other off-sets were included to control for unbalanced data and non-independence (e.g., some individuals might sound very much alike and we included multiple pairs of calls from the same two individuals).

Model II: decay over time within recording

To test how stable the individual signature is within a recording we subsetted the pairwise dis-similarities to only include comparisons between calls from the same individual and recording. We included log10 of the time (in minutes) between the calls. We then used a multi-level model with a global intercept, a slope for the time-effect, off-sets for the intercept pair of individuals from which the calls came, pair of recordings from which the calls came and each call. We also included off-sets on the slope for individual. The full model structure was as follows:

$$
\begin{aligned}
\text{accoustic distance} &\sim \text{normal}(\mu,\ \sigma) \\
\mu_n &= \bar{\alpha} + \alpha_{\text{ind[n]}} + \alpha_{\text{rec[n]}} + \alpha_{\text{call i[n]}} + \alpha_{\text{call i[n]}} \\
&\quad (\bar{\beta} + \beta_{\text{ind[n]}} + \beta_{\text{rec[n]}}) * \text{time} \\
\bar{\alpha} &\sim \text{normal}(0,\ 0.5) \\
\alpha_{\text{ind}} &\sim \text{normal}(0,\ \sigma_{\text{ind}}) \\
\alpha_{\text{rec}} &\sim \text{normal}(0,\ \sigma_{\text{rec}}) \\
\alpha_{\text{call}} &\sim \text{normal}(0,\ \sigma_{\text{call}}) \\
\bar{\beta} &\sim \text{normal}(0,\ 0.3) \\
\beta_{\text{ind}} &\sim \text{normal}(0,\ \xi_{\text{ind}}) \\
\beta_{\text{rec}} &\sim \text{normal}(0,\ \xi_{\text{rec}}) \\
\sigma,\ \sigma_{\text{ind}},\ \sigma_{\text{rec}},\ \xi_{\text{ind}},\ \xi_{\text{rec}} &\sim \text{exponential}(2) \\
\sigma_{\text{call}} &\sim \text{exponential}(3)
\end{aligned}
$$

The model was fitted in the same way as model I.

Model III: decay over days

To test how stable the individual signature is across days we subsetted the pairwise dis-similarities to only include comparisons between calls from the same individual and year, but from different recording. We also included the difference in days between the calls. We then used a multi-level model with a global intercept, a slope for the time-effect, off-sets for the intercept for individual pair, recording and call, and off-sets on the slope for individual and recording. The full model structure was as follows:

$$
\begin{aligned}
\text{accoustic distance} &\sim \text{normal}(\mu,\ \sigma) \\
\mu_n &= \bar{\alpha} + \alpha_{\text{ind pair[n]}} + \alpha_{\text{rec pair[n]}} + \alpha_{\text{call i[n]}} + \alpha_{\text{call i[n]}} \\
&\quad (\bar{\beta} + \beta_{\text{ind[n]}}) * \text{time} \\
\bar{\alpha} &\sim \text{normal}(0,\ 0.5) \\
\alpha_{\text{ind pair}} &\sim \text{normal}(0,\ \sigma_{\text{ind pair}}) \\
\alpha_{\text{rec pair}} &\sim \text{normal}(0,\ \sigma_{\text{rec pair}}) \\
\alpha_{\text{call}} &\sim \text{normal}(0,\ \sigma_{\text{call}}) \\
\bar{\beta} &\sim \text{normal}(0,\ 0.5) \\
\beta_{\text{ind}} &\sim \text{normal}(0,\ \xi_{\text{ind}}) \\
\sigma,\ \sigma_{\text{ind pair}},\ \sigma_{\text{rec pair}},\ \sigma_{\text{call}},\ \xi_{\text{ind}} &\sim \text{exponential}(2)
\end{aligned}
$$

The model was fitted in the same way as model I.

Model IV: difference between years

To test how stable the individual signature is across years we subsetted the pairwise dis-similarities to only include comparisons between calls from the same individual, but from different recording. We also included whether or not the recordings came from the same year. We then used a multi-level model with a global intercept, off-sets for: same vs different year, the individual, recording pair and call. The full model structure was as follows:

$$
\begin{aligned}
\text{accoustic distance} &\sim \text{normal}(\mu,\ \sigma) \\
\mu_n &= \bar{\alpha} + \alpha_{\text{same year[n]}} + \alpha_{\text{ind[n]}} + \\
&\quad \alpha_{\text{rec pair[n]}} + \alpha_{\text{call i[n]}} + \alpha_{\text{call i[n]}} \\
\bar{\alpha} &\sim \text{normal}(0,\ 0.25) \\
\alpha_{\text{same year}} &\sim \text{normal}(0,\ \sigma_{\text{same year}}) \\
\alpha_{\text{ind}} &\sim \text{normal}(0,\ \sigma_{\text{ind}}) \\
\alpha_{\text{rec pair}} &\sim \text{normal}(0,\ \sigma_{\text{rec pair}}) \\
\alpha_{\text{call}} &\sim \text{normal}(0,\ \sigma_{\text{call}}) \\
\sigma,\ \sigma_{\text{same year}},\ \sigma_{\text{ind}},\ \sigma_{\text{rec pair}},\ \sigma_{\text{call}} &\sim \text{exponential}(3)
\end{aligned}
$$

The model was fitted in the same way as model I.

pDFA

To test if there are features that contain an individual signal across call types we used mel frequency cepstral coefficients (MFCC). More specifically we used the function *melfcc* from the package *warbleR* (Araya-Salas and Smith-Vidaurre 2017) with the settings wintime = 512/44100, hoptime = 50/44100 and minfreq = 300 to retrieve the first ten cepstral coefficient traces. For each trace we saved the mean and standard deviation. These 20 measures per call were then used for the pDFA.

We wrote a custom R script to run the pDFA. We ran 100 iterations with the following steps:

- for each individual we randomly selected recordings until the set of recordings contained enough calls for testing (N_test)
- we then selected the other recordings for training
- if there were not enough recordings to fullfill both N_test and N_train, we did not include this individual
- for the individuals with enough data, we randomly selected N_test calls from the testing recordings and N_train calls from the training recordings
- we then used the standardised MFCC values to train an LDA (using the function *lda* from the package *MASS* (Venables and Ripley 2002))
- this LDA was then used to predict the IDs of the test data, the resulting proportion of correct classification was the trained score
- we repeated the previous two steps with both training and testing IDs randomised within the set and retrieved the random score
- for each iteration the difference between the trained and random score was reported

Supplemental Results

pDFA

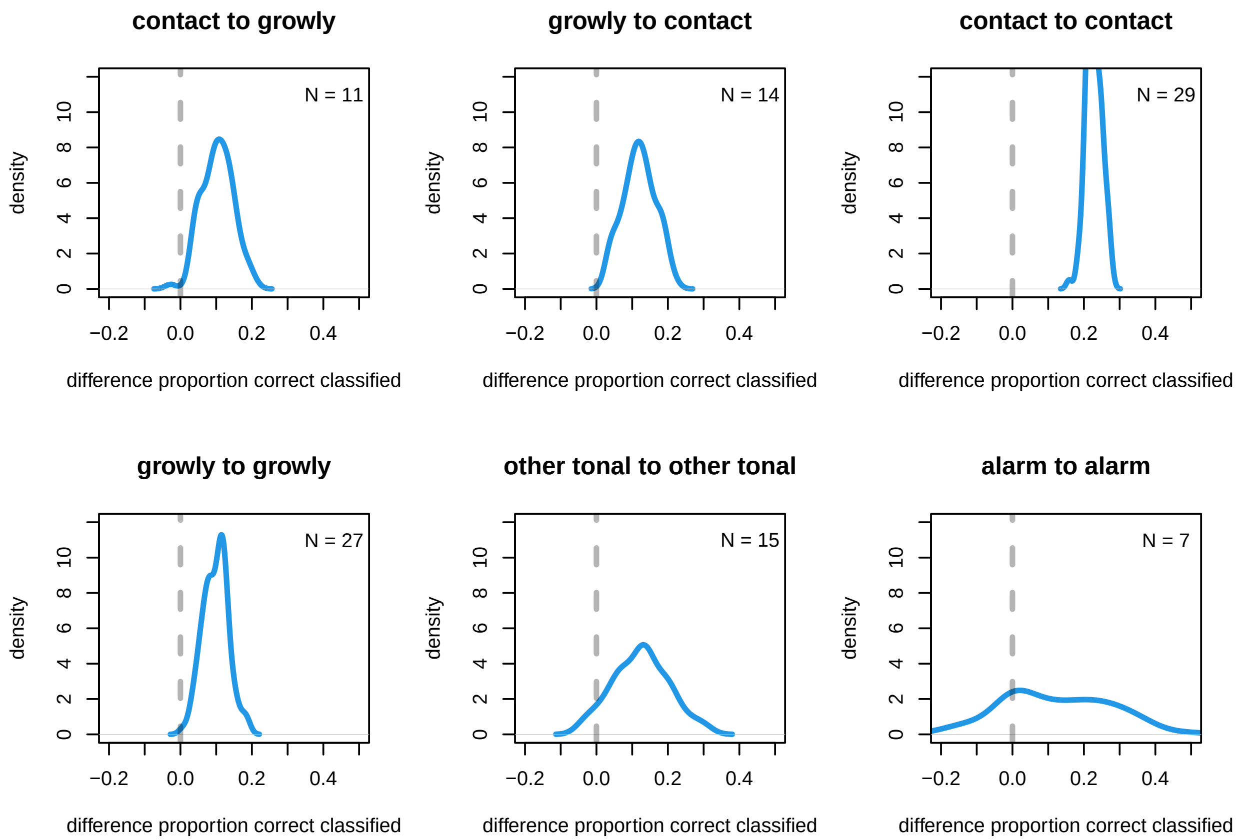

Figure S1: pDFA results. Densities of the differences between the trained and random DFAs. N is number of individals for which there was enough data available.

References

Araya-Salas, M., and G. Smith-Vidaurre. 2017. "warbleR: An r Package to Streamline Analysis of Animal Acoustic Signals." *Methods in Ecology and Evolution*. https://dx.doi.org/10.1111/2041-210X.12624.

Gabry, Jonah, and Rok Češnovar. 2021. *Cmdstanr: R Interface to 'CmdStan'*.

Gelman, Andrew, Daniel Lee, and Jiqiang Guo. 2015. "Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization." *Journal of Educational and Behavioral Statistics* 40 (5): 530–43.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with s*. Fourth. New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/.