

# Voice paper

May 2022

## Contents

<b>1</b>	<b>Supplemental Methods</b>	<b>1</b>
1.1	Call types . . . . .	1
1.2	Dynamic time warping (DTW) . . . . .	2
1.3	Statistical analysis . . . . .	2
<b>2</b>	<b>Supplemental Results</b>	<b>6</b>
2.1	Model IV: difference between years . . . . .	6
2.2	pDFA . . . . .	7
	<b>References</b>	<b>11</b>

## 1 Supplemental Methods

### 1.1 Call types

Call types were manually determined from the spectrogram and audio in Raven Lite (K. Lisa Yang Center for Conservation Bioacoustics 2016). We included X types in the main text (see also figure SX):

- contact: tonal vocalisation component and at least three frequency modulation peaks
- tja: tonal vocalisation with a single up-sweep component; sometimes one or two very short amplitude modulated components are present before or after the main component
- trruup: vocalisation that consists of at least three amplitude modulated components at the start, followed by a frequency modulated tonal component
- alarm: loud amplitude modulated tonal vocalisation with at least five clear u-shaped components
- growl: medium loud amplitude modulated vocalisation with at least five components without clear tonality

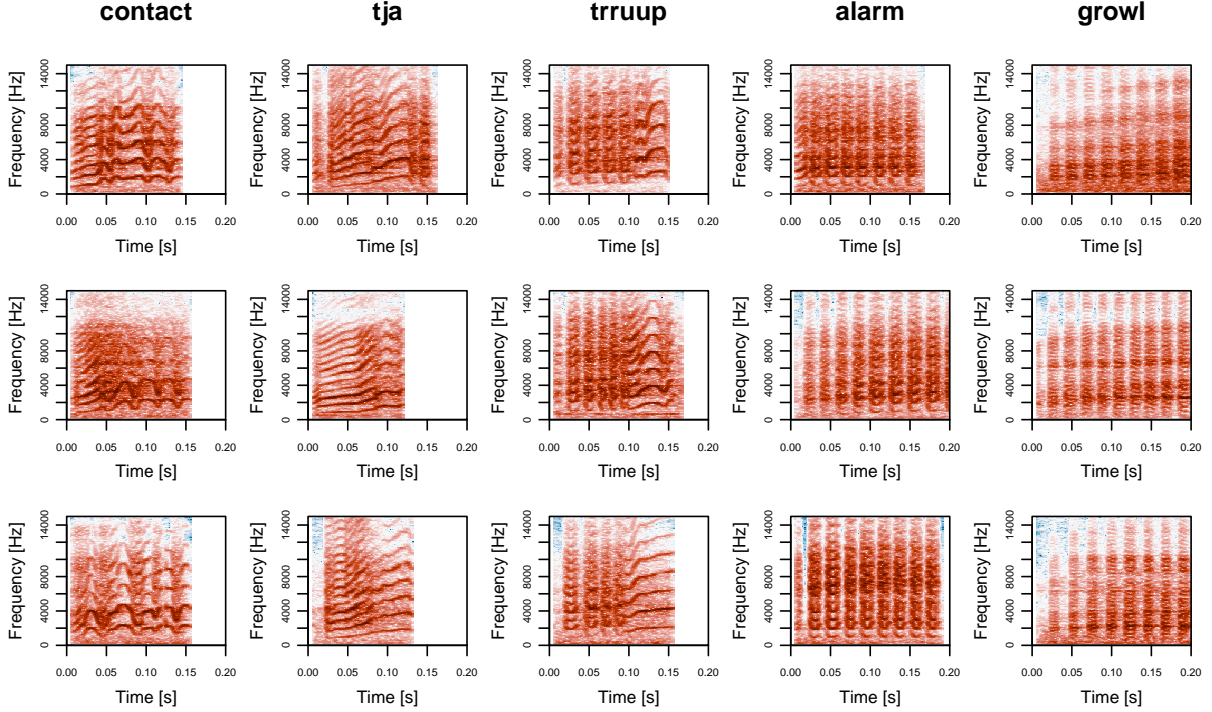


Figure S1: Example spectrograms of the call types.

## 1.2 Dynamic time warping (DTW)

For tonal calls the fundamental frequency was traced in *Luscinia* ([lachlan2007luscini](#)?). For non-tonal calls only the start and end times were determined manually. Traces were then exported as CSV and imported into R (R Core Team 2021) for further analysis. Traces were smoothed using two steps. The first step filled gaps within a call by connecting the fundamental frequency at both sides of the gap with a straight line. To get rid of small measurement jitter, we used `smooth.spline` from *stats* with `spar = 0.1`.

[ADD DTW, MFCCCC and SPECAN]

## 1.3 Statistical analysis

### 1.3.1 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:

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

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) 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).

### 1.3.2 Model II: decay over time within recording

To test how stable the individual signature is within a recording we subsetting 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:

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

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

### 1.3.3 Model III: decay over days

To test how stable the individual signature is across days we subsetting the pairwise dis-similarities to only include comparisons between calls from the same individual and year, but from different recordings. 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. The full model structure was as follows:

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

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

#### 1.3.4 Model IV: difference between years

To test how stable the individual signature is across years we subsetting 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:

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

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

#### 1.3.5 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 ( $N_{\text{test}}$ )
- we then selected the other recordings for training
- if there were not enough recordings to fulfill both  $N_{\text{test}}$  and  $N_{\text{train}}$ , we did not include this individual

- for the individuals with enough data, we randomly selected  $N_{\text{test}}$  calls from the testing recordings and  $N_{\text{train}}$  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

## 2 Supplemental Results

### 2.1 Model IV: difference between years

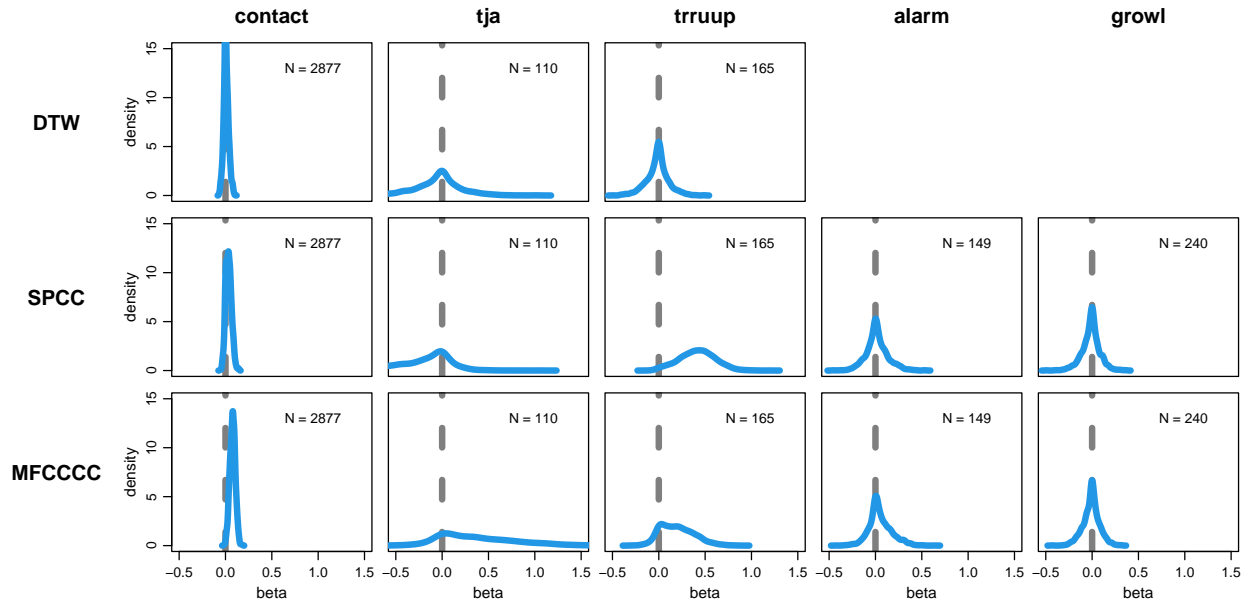


Figure S2: Parameter estimates (posterior distributions) for the Bayesian multilevel model with difference between recordings from the same vs different years. Greater values mean more within year similarity. Models were run separately for the main call types (columns) and methods (rows; DTW - dynamic time warping, SPCC - spectrographic cross correlation, MFCC - mel frequency cepstral coefficient analysis). N represents the number of calls included in the model.

## 2.2 pDFA

For the results presented in the paper there was no overlap with zero for any of the pDFAs. For full distributions and sample sizes see Figure 3.

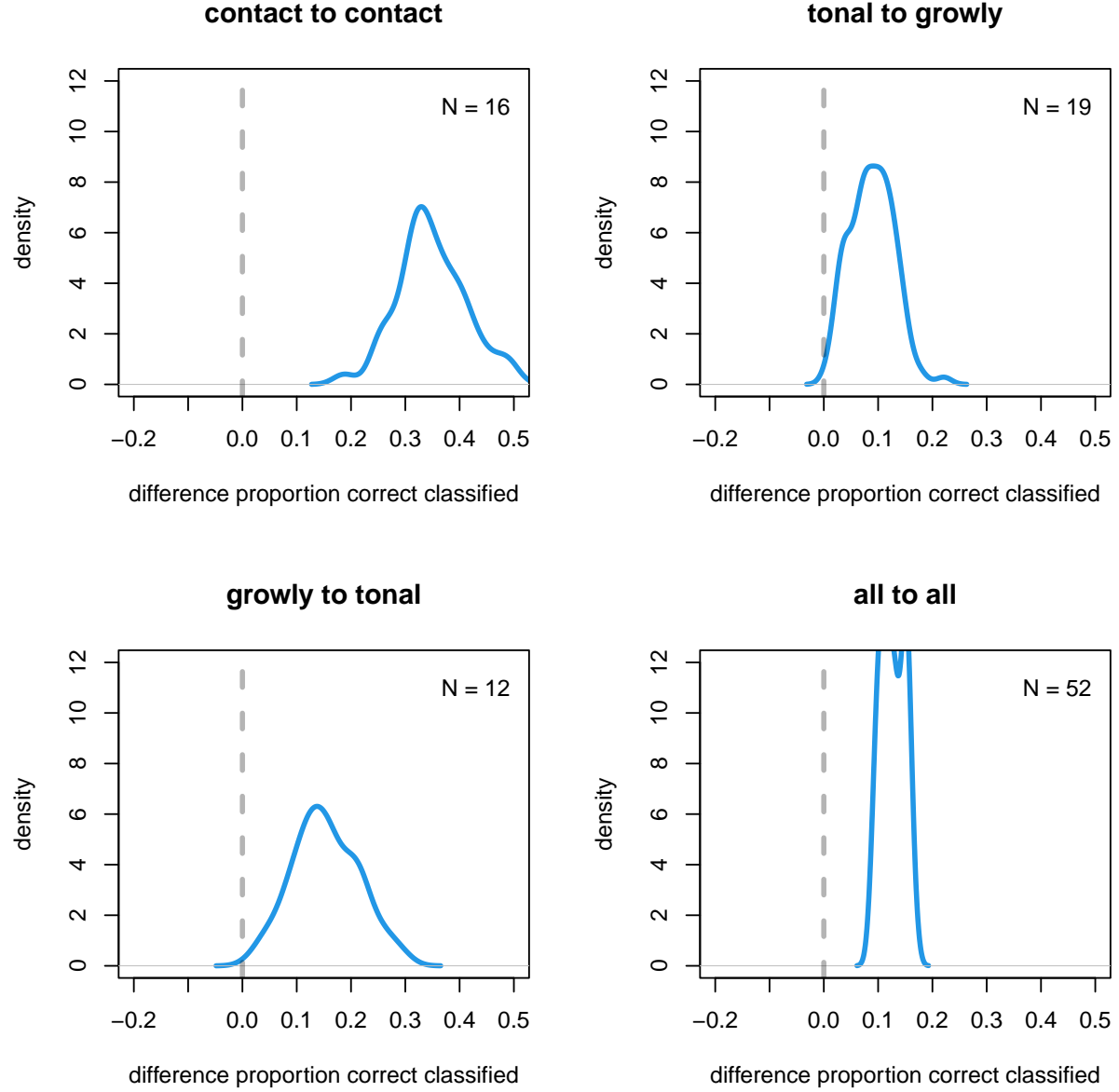


Figure S3: pDFA results. Densities of the differences between the trained and random DFAs. N is number of individuals for which there was enough data available.

When sub-setting the data to only include females from the square the results for growly to tonal fell towards chance-level, but all other results remained stable with overlap with zero less than 2%. For full distributions and sample sizes see Figure 4.

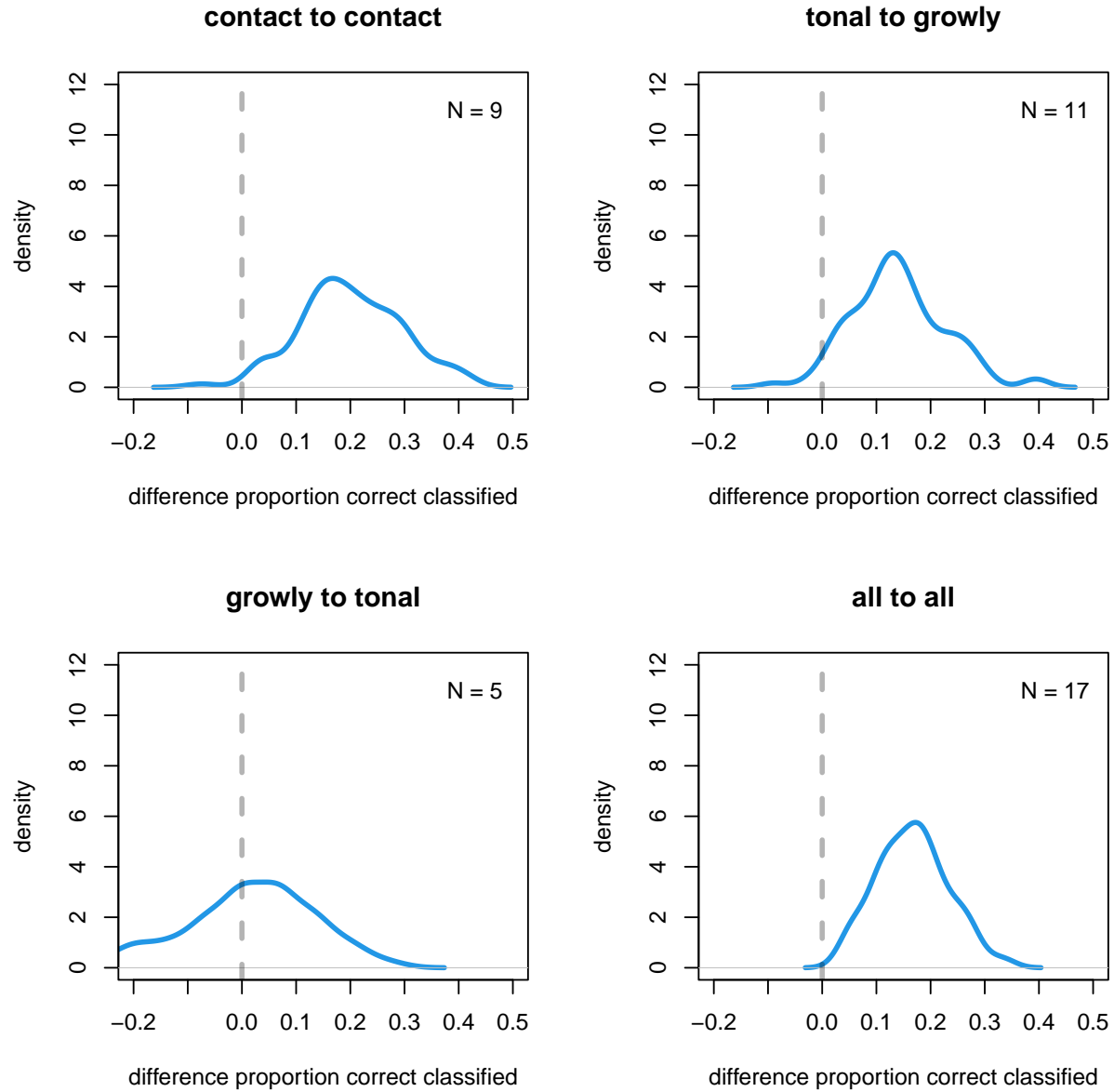


Figure S4: pDFA results for females from the square. Densities of the differences between the trained and random DFAs. N is number of individuals for which there was enough data available.



When sub-setting the data to only include males from the square the results for both cross-call type models fell towards chance-level. For full distributions and sample sizes see Figure 5.

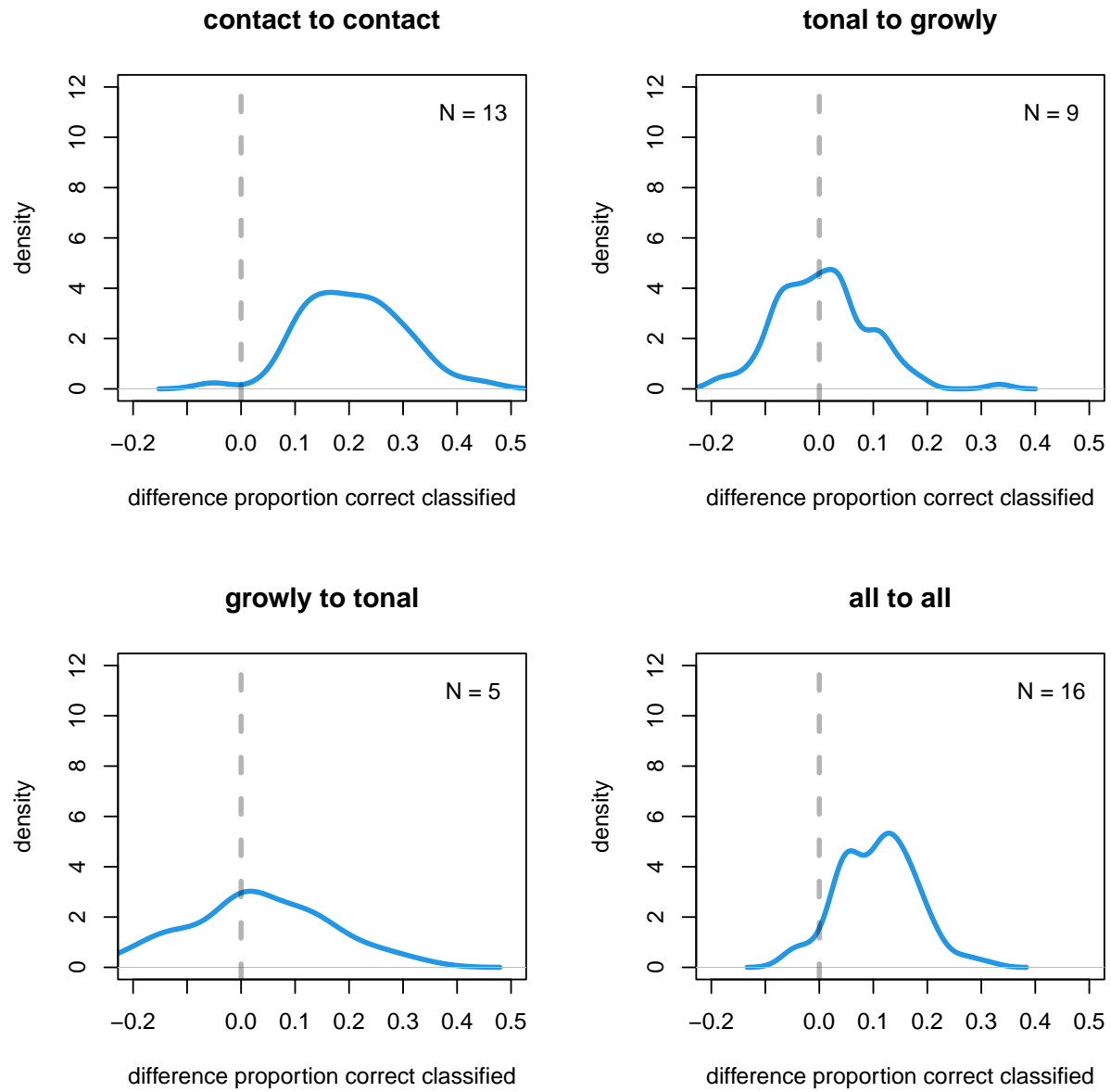


Figure S5: pDFA results for males from the square. Densities of the differences between the trained and random DFAs. N is number of individuals for which there was enough data available.

When sub-setting the data to only include females from the square the results for the cross call type models remained stable but overlap with zero increased to 6-7%. For full distributions and sample sizes see Figure 6.

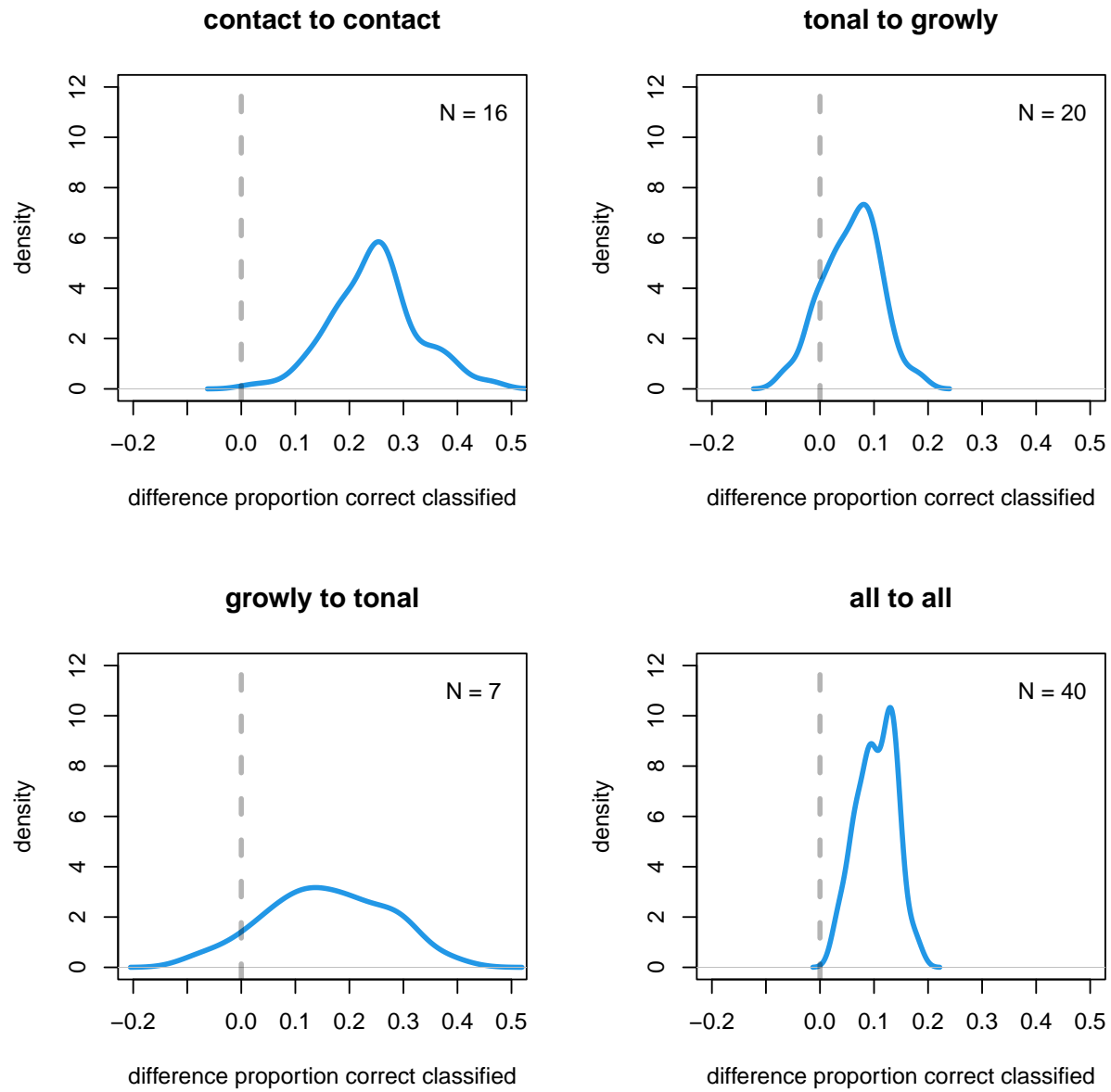


Figure S6: pDFA results with randomisation withing square and park only. Densities of the differences between the trained and random DFAs. N is number of individuals for which there was enough data available.

## 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.
- K. Lisa Yang Center for Conservation Bioacoustics. 2016. *Raven Lite: Interactive Sound Analysis Software* (version 2.0.1). <https://ravensoundsoftware.com>.
- 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/>.