A universal symphony: coral reef fish calls exhibit consistent acoustic characteristics across different bioregions

Lucille Chapuis*, Tzu-Hao Lin, Ben Williams, Timothy A. C. Lamont, Rucha Karkarey, Gabriela G. Nava-Martínez, Aya Maryam Rahil Naseem, Andrew N. Radford, Stephen D. Simpson

Supplementary Method 1: comparison of acoustic features in True Positive and False Negative detections

To evaluate whether the detection model selectively captured a subset of fish calls and potentially biased the diversity analysis, we compared some acoustic features of true positive (TP) detections (correctly identified calls) and false negatives (FN) (annotated calls missed by the model). We extracted four parameters from the manual annotations: call duration, low frequency (Hz), high frequency (Hz) and sginal-to-noise ratio (SNR). Duration was computed as the difference between the end and start time of each call. Frequency parameters were taken directly from the 'Low Freq (Hz)' and 'High Freq (Hz)' columns of the annotation files (standard Raven Pro format). For each call, we extracted the waveform segment corresponding to the call's start and end time, and a short preceding window (50 ms) of ambient sound to represent background noise. The SNR was then calculated as the ratio of the root-mean-square (RMS) amplitude of the signal to that of the noise, expressed in decibels (dB). This approach assumes that the background noise immediately preceding a call is representative of the local ambient noise environment.

Mann–Whitney U tests were performed to assess whether the distributions of these parameters differed significantly between TP and FN calls. The results showed no significant difference in low frequency ($U = 663068.5$, $p = 0.89$) or high frequency ($U = 657108.0$, $p = 0.50$) distributions between groups. Call duration showed a minor trend toward shorter durations in false negatives, but this was not statistically significant ($U = 599040.5$, 0.09). In contrast, the signal-to-noise ratio (SNR) revealed a highly significant difference ($U = 12346112.0$, $p < 0.0001$), indicating that missed calls tended to have lower SNR values compared to detected ones (Figure S1).

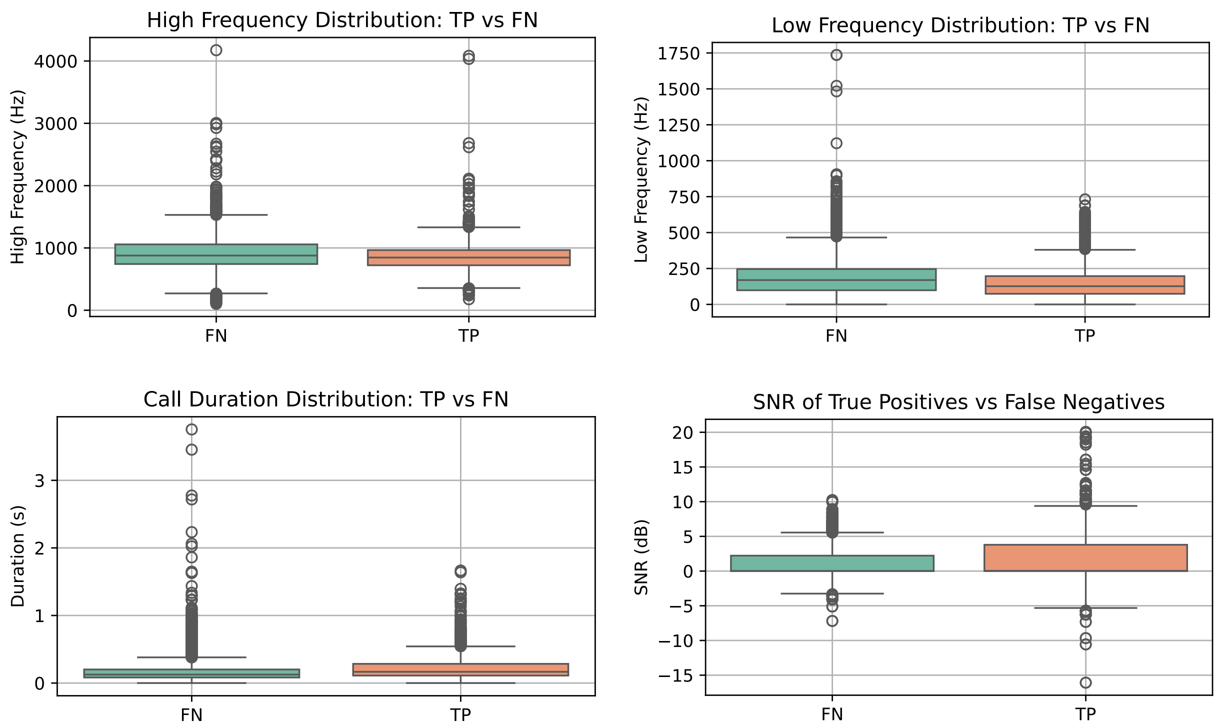

Figure S1: Comparison of acoustic parameters between true positive (TP) detections and false negatives (FN). Boxplots show the distribution of (A) call duration (s), (B) low frequency (Hz), (C) high frequency (Hz), and (D) signal-to-noise ratio (SNR, dB) for calls correctly detected by the model (TP) and calls missed by the model (FN). No significant differences were observed in frequency parameters or call duration. However, SNR was significantly lower for FN calls (Mann–Whitney U test, $p < 0.0001$).

These findings suggest that the main limitation of the detection model relates to amplitude or background noise conditions, rather than spectral or temporal structure of the calls. Thus, although the model underperformed in detecting lower SNR events, it did not systematically exclude calls with specific spectral characteristics or durations. The results of the diversity analysis presented in this study therefore likely reflect a representative sample of the broader fish acoustic community.

Supplementary Method 2: Comparison of retained and rejected fish calls

To assess potential bias introduced by the geometric morphometric (GMM) alignment step, we compared the acoustic properties of detected fish calls that were successfully aligned ("retained") versus those that failed alignment ("rejected"). For each of the six datasets, we identified call files that were present in the raw detection folder but absent from the aligned call folder. We extracted two basic acoustic parameters for each call using Python: (1) call duration, and (2) dominant frequency, identified as the frequency with the highest spectral amplitude. We then visualised and compared the distributions of these features between retained and rejected calls across all datasets. Summary statistics (mean, standard deviation, and median) were calculated for each group (Table S1).

Table S1: Summary statistics for each dataset (location) on the duration and dominant frequency calculated for rejected versus kept calls for the GMM analysis.

		Duration			Dominant frequency		
Dataset	label	mean	std	median	mean	std	median
Australia	kept	0.079	0.019	0.073	512	630	337
	rejected	0.077	0.019	0.073	601	795	311
French Polynesia	kept	0.081	0.037	0.073	841	745	554
	rejected	0.080	0.038	0.073	688	686	513
India	kept	0.081	0.020	0.073	409	310	311
	rejected	0.079	0.031	0.073	431	287	364
Indonesia	kept	0.097	0.071	0.074	442	655	256
	rejected	0.091	0.058	0.074	481	775	175
Maldives	kept	0.101	0.077	0.074	842	788	567
	rejected	0.087	0.048	0.073	792	741	554
Mexico	kept	0.097	0.052	0.074	599	810	243
	rejected	0.084	0.031	0.073	469	686	216

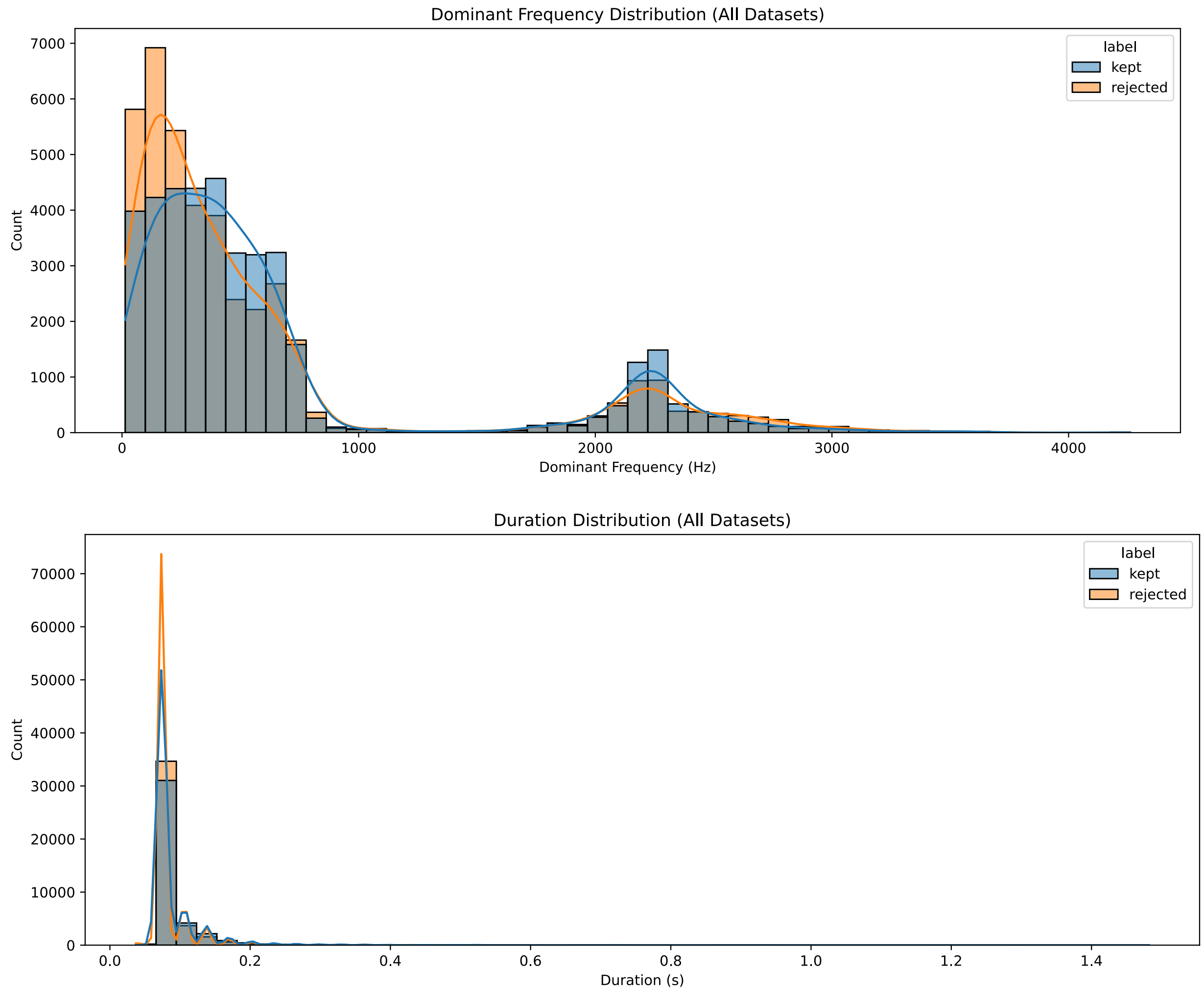

Figure S2: Density plot of the (A) dominant frequency and (B) duration distribution of the rejected versus kept calls for the GMM analysis.

Across datasets, the distributions of call durations were broadly similar between retained and rejected groups, with median durations consistently around 0.074 s. Mean dominant frequency varied more, but no consistent directional bias was observed. We note a slight peak in dominant frequency below 300 Hz among the rejected calls, which may reflect a higher influence of low-frequency background noise. These lower-frequency calls are more susceptible to masking by ambient or anthropogenic noise sources (e.g., surf, distant boat engines), potentially reducing their signal-to-noise ratio and leading to failed alignment in the GMM preprocessing pipeline. Standard deviations were generally large across both groups, reflecting high within-group variability. These findings suggest that the alignment failures were not systematically biased toward a particular frequency or duration range, supporting the use of the aligned subset as a representative sample for shape-based analysis.