Efficient Auditory Coding for Bat Vocalizations

Testing Auditory Kernel Effciency on Rhinolophus Affnis Calls

Supervisor: Dimme de Groot   Author: Aleksandra Savova   Responsible Professor: Jorge Martinez Castaneda

EEMCS, Delft University of Technology

Introduction into Auditory Kernels

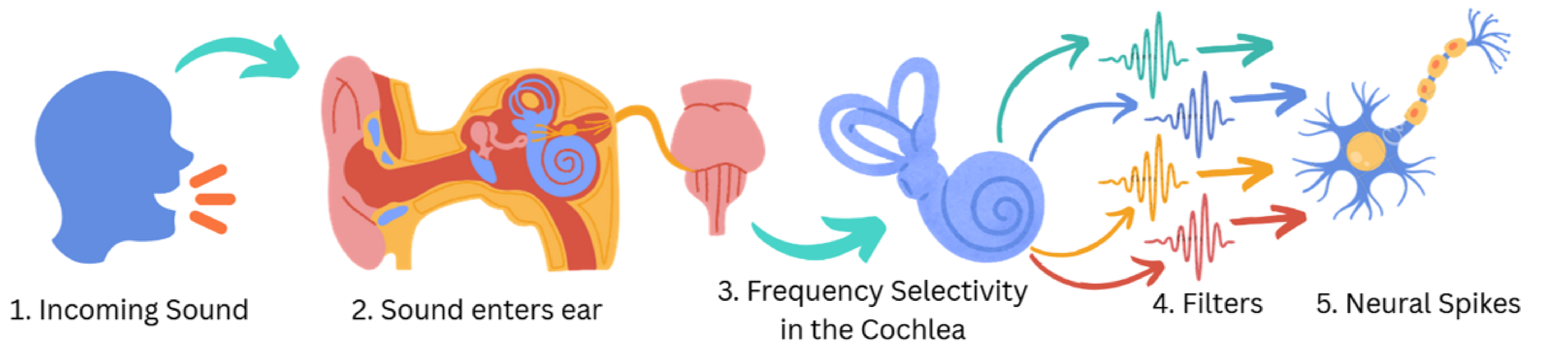

Figure 1. **Human auditory processing.** Schema of the early auditory system illustrating the efficient coding hypothesis.

Sparse Coding. Decompose signals $x(t)$ into a set of discrete acoustic elements, called auditory kernels, $\phi_{\gamma_k}(t)$, each scaled by $a_k \in \mathbb{R}$ and characterized by a set of parameters $\gamma_k = (\tau_k, \omega_k, s_k) \in \Gamma$, specifying the time shift τ_k, center frequency ω_k, and scale s_k [1]. The decomposition is given by:

$$x(t) \approx \sum_{k=1}^{K} a_k \phi_{\gamma_k}(t), \quad K \ll |\Gamma|$$

where K is the number of active atoms and Γ the set of all possible atom parameters.

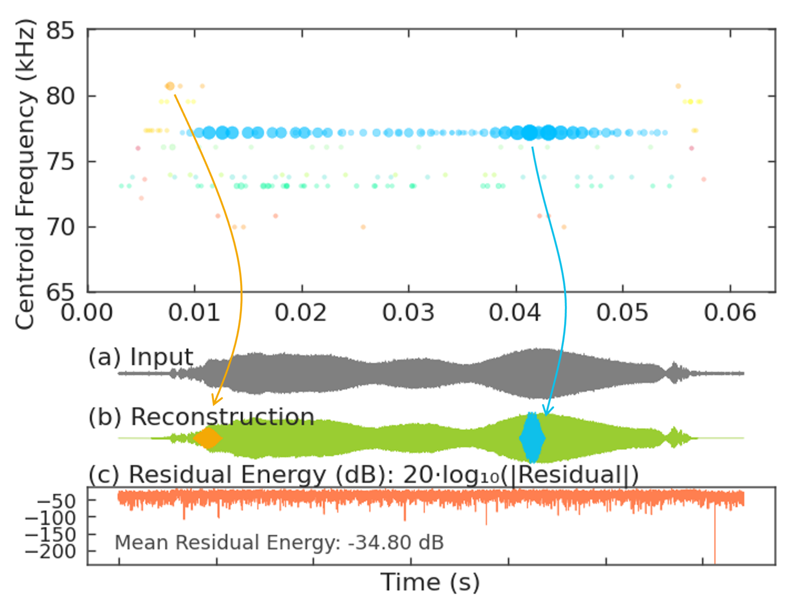

Figure 2. **Sparse coding decomposition of a bat echolocation call.** Reconstruction with 200 spikes shown over time and centroid frequency with MSE 1.918×10^{-4} and SNR 20.62 dB.

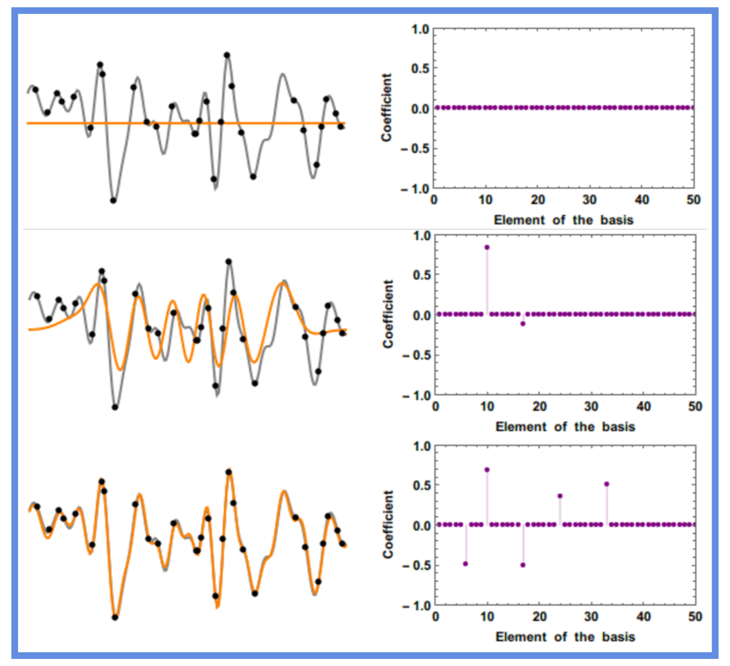

Figure 3. **Matching pursuit decomposition.** Iterative signal approximation (orange line) with kernel spikes (purple).

The ChiroVox Dataset and Bat Echolocation Calls

Echolocation as species-specific auditory trait of Rhinolophus affinis and Rhinolophus pearsonii.

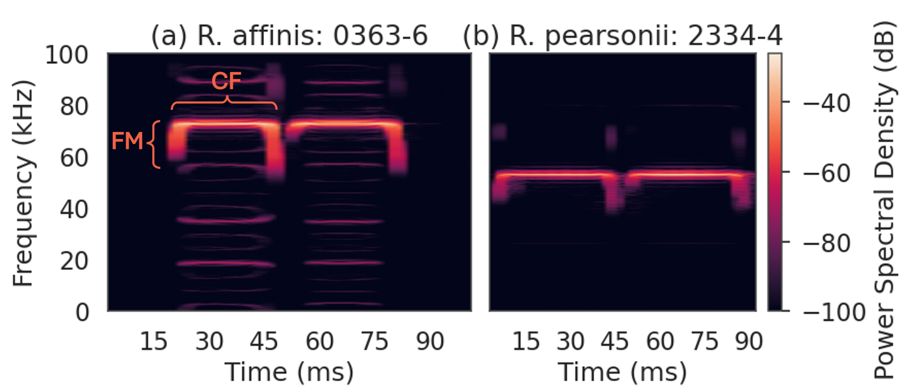

Figure 4. **Echolocation call spectrograms of two Rhinolophidae species.** (a) *R. affinis* calls at 72.9 kHz, with visible secondary harmonics (faint horizontal lines). (b) *R. pearsonii* calls at 53.2 kHz. [2]

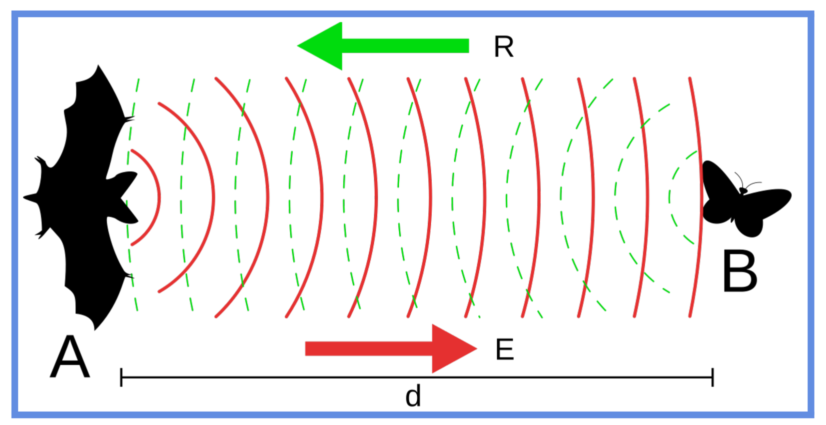

Figure 5. **Bat echolocation diagram.** A: Bat, B: Prey, E (red): Echolocation call emitted by the bat, R (green): Response or echo received by the bat.

Research Questions

RQ 1. What spectral characteristics define the structure of the kernels learned through the sparse coding of bat vocalizations?

RQ 2. To what extent do sparse representations achieve greater coding efficiency compared to traditional signal representation methods, such as Fourier and wavelet transforms?

RQ 3. To what degree do the learned kernels show functional specialization, with clusters of similar activation profiles encoding specific variations in bat calls?

RQ 4. To what extent do the learned representations exhibit sparsity, with a high prevalence of inactive (near-zero) coefficients across the kernel dictionary?

Frequency Distribution Analysis and Denoising

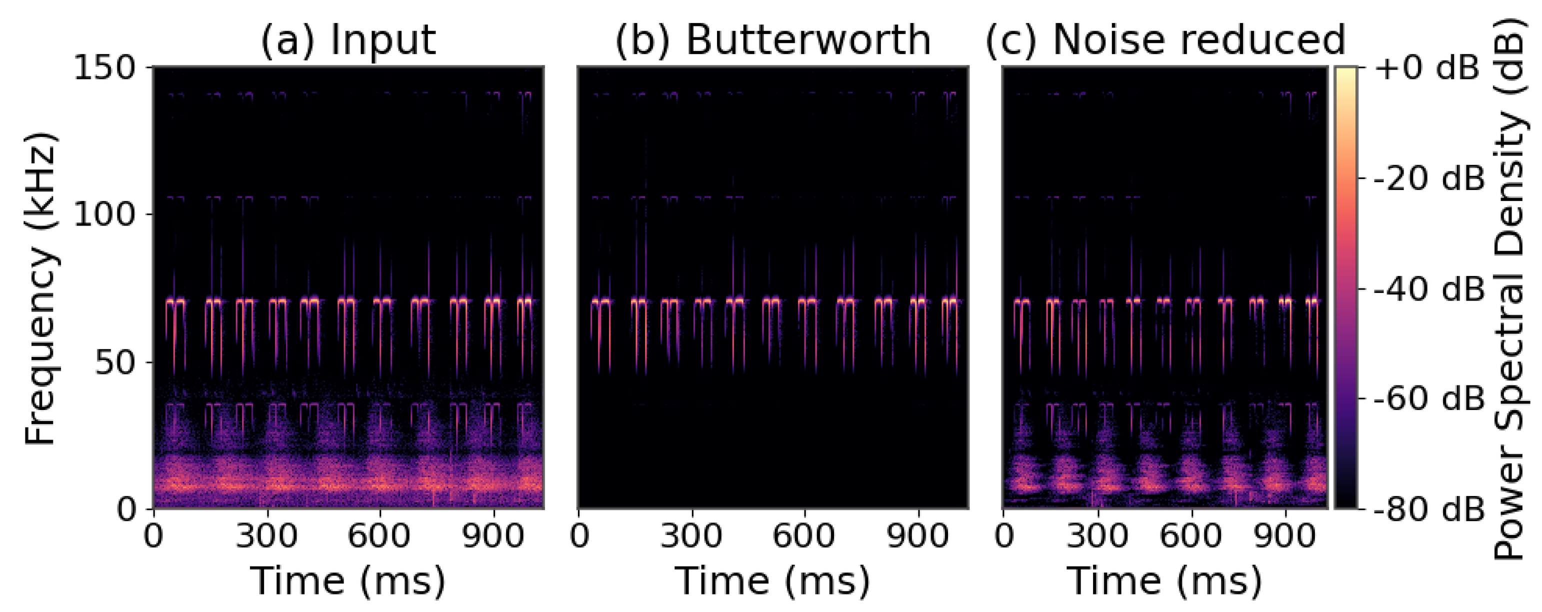

Figure 6. **Denoising performance on *R. affinis* echolocation call.** Recording with 500 kHz sample rate, spectral peak at 70.56 kHz, and Butterworth cutoff at 49.39 kHz. Python `noisereduce` library shows poor performance (c).

Energy-Based Bat Call Detection

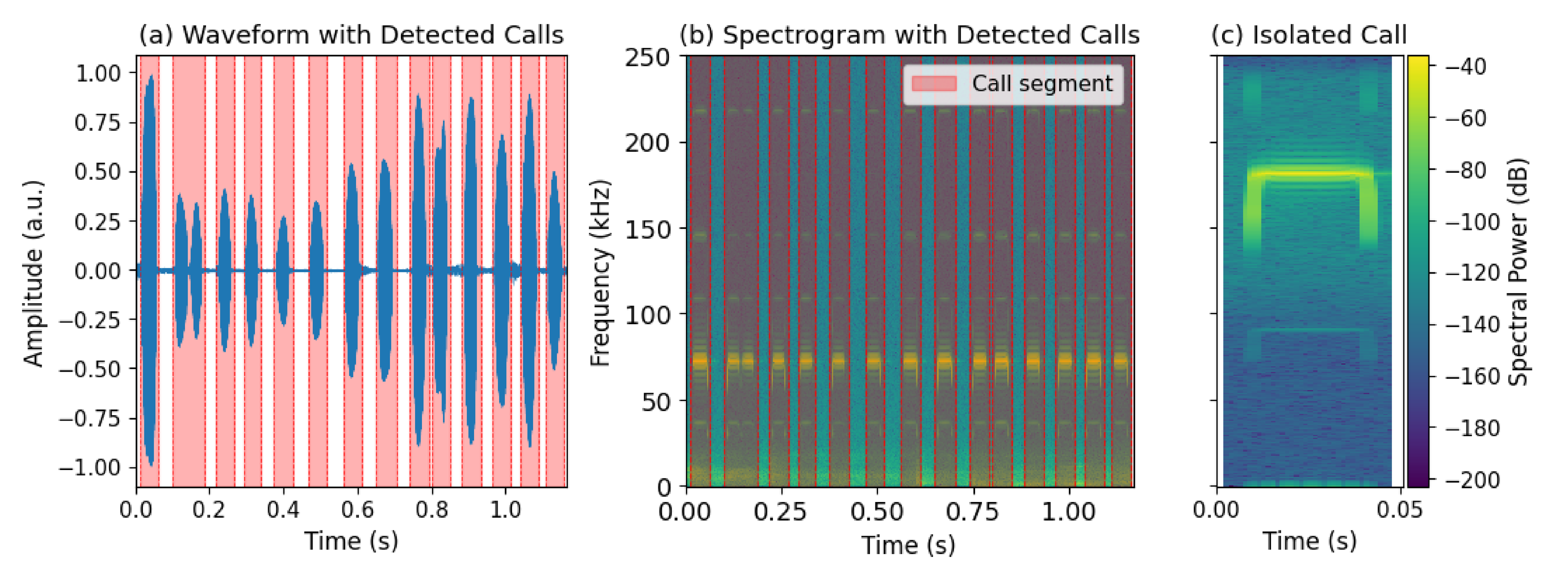

Figure 7. **Energy-based call detection for *R. affinis*.** [3] Calls are identified where energy exceeds a 3 dB threshold and padded with 5 ms to make sure to capture FM sweeps.

References

[1] E. C. Smith and M. S. Lewicki, "Efficient auditory coding," *Nature*, 2006.

[2] T. Görföl, J. C.-C. Huang, G. Csorba, D. Gyfrössy, P. Estók, T. Kingston, and et al., "Chirovox: a public library of bat calls," *PeerJ*, vol. 10, p. e12445, 2022.

[3] M. D. Skowronski and M. B. Fenton, "Quantifying bat call detection performance of humans and machines," *Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 513–521, 2009.

Spectral Characteristics (RQ 1) and Efficiency (RQ 2)

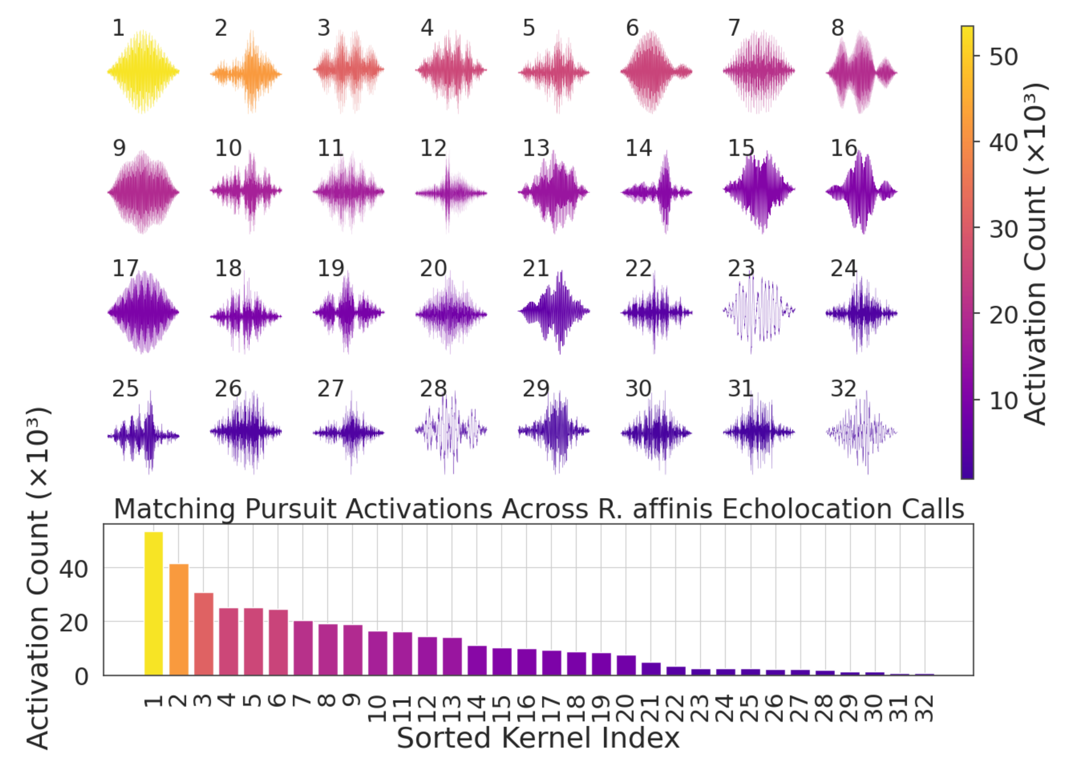

Figure 8. **Kernel activation in *R. affinis* echolocation.** 32 kernels (400 samples each) trained on echolocation with 10,000 gradient ascent iterations, sorted by activation in 70-spike reconstruction.

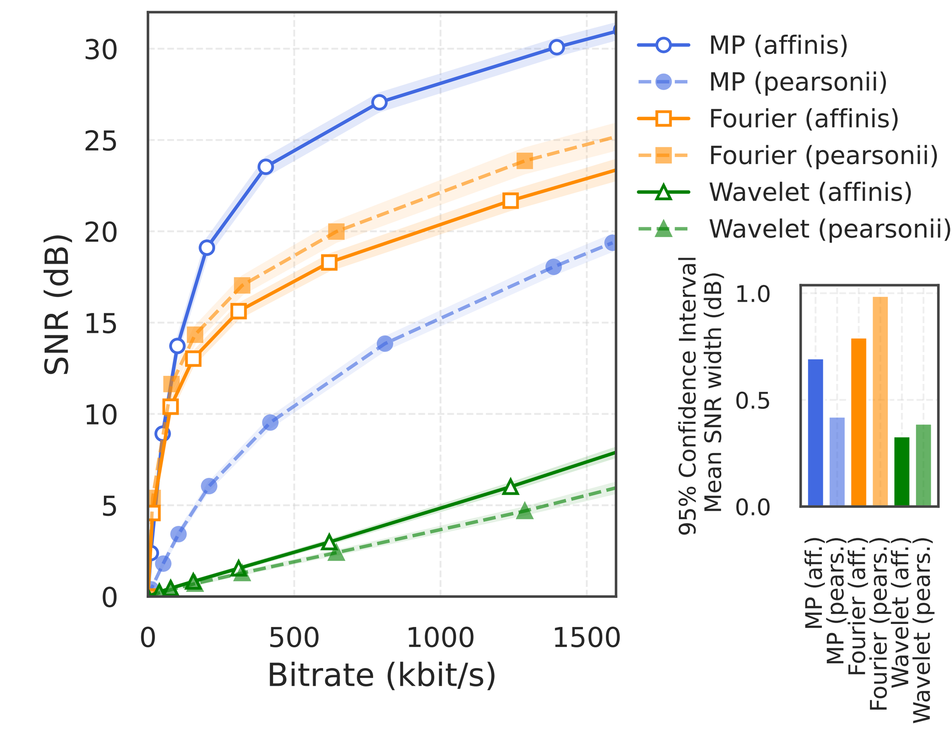

Figure 9. **Fidelity rate of Matching Pursuit, Fourier, and wavelet coding for *R. affinis* and *R. pearsonii* calls.** Shaded regions denote tight (<1 dB) 95% confidence intervals.

Kernel Functional Specialization (RQ 3)

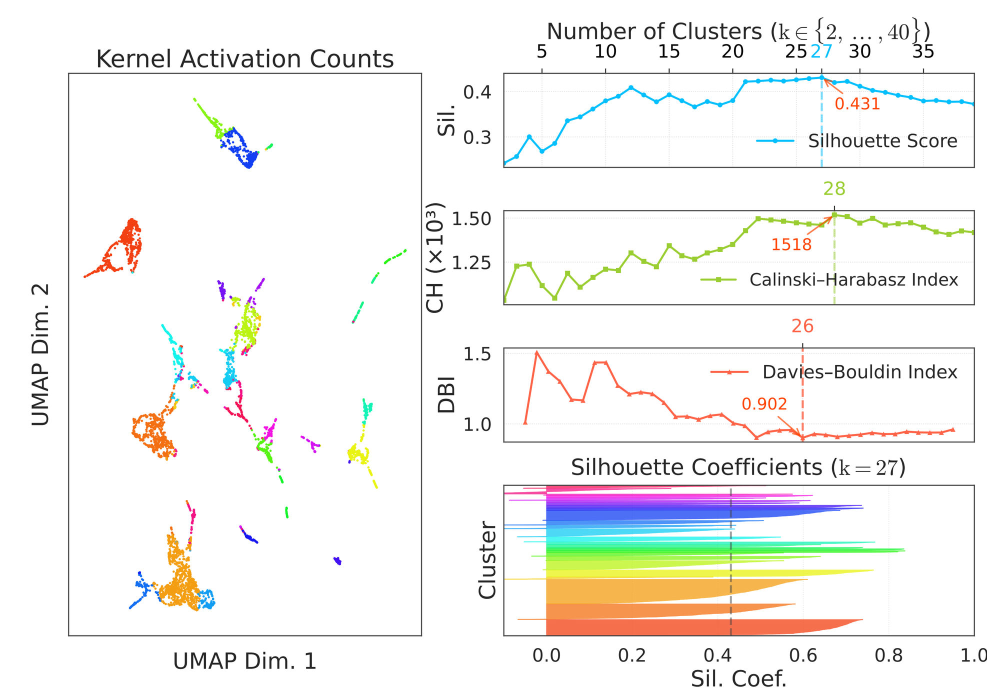

Figure 10. **Clustering metrics and UMAP projection of kernel activations.** Kernel activations clustering with reconstruction depth and cluster number chosen using the Silhouette score (Sil.), Calinski–Harabasz Index (CHI), and Davies–Bouldin Index (DBI).

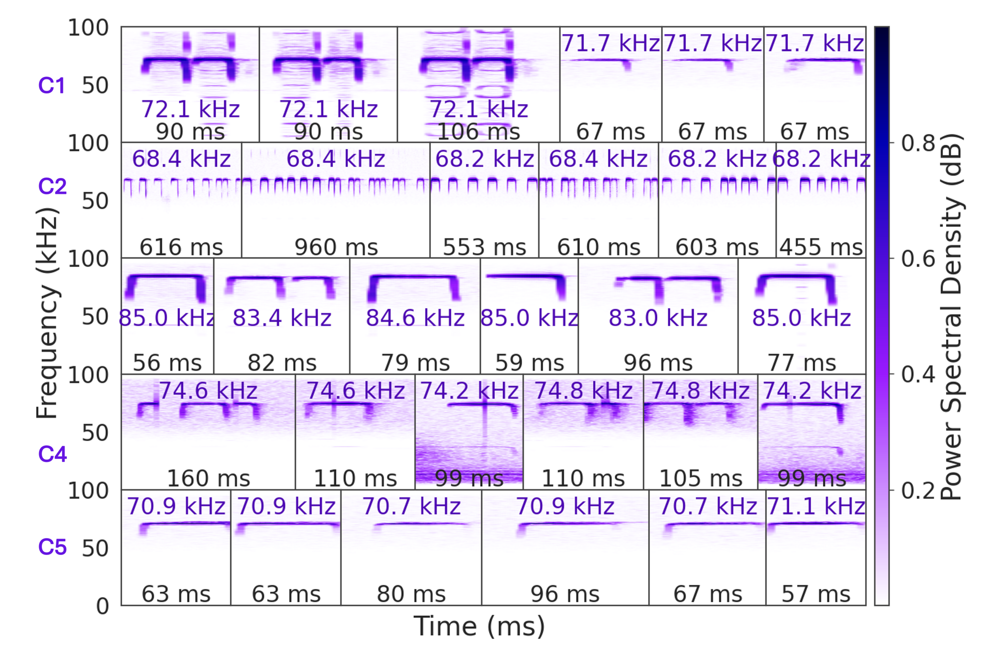

Figure 11. **Auditory kernel-based clustering captures call diversity.** Five selected clusters (C1–C5) with six central calls with scaled durations per row.

Activation Sparsity (RQ 4)

Metric	200	2400
Gini (aff.)	0.985 ± 0.008	0.997 ± 0.001
Gini (pea.)	0.994 ± 0.004	0.998 ± 0.001
Hoyer (aff.)	0.960 ± 0.028	0.981 ± 0.011
Hoyer (pea.)	0.996 ± 0.013	0.993 ± 0.006
PQ (aff.)	$(3.50 \pm 0.31) \times 10^{-3}$	$(2.3 \pm 0.1) \times 10^{-4}$
PQ (pea.)	$(6.9 \pm 1.5) \times 10^{-4}$	$(1.0 \pm 0.1) \times 10^{-4}$

Table 1. **Sparsity metrics.** Mean ± SD of Gini, Hoyer, and PQ over 1000 recordings for *R. affinis* and *R. pearsonii.*

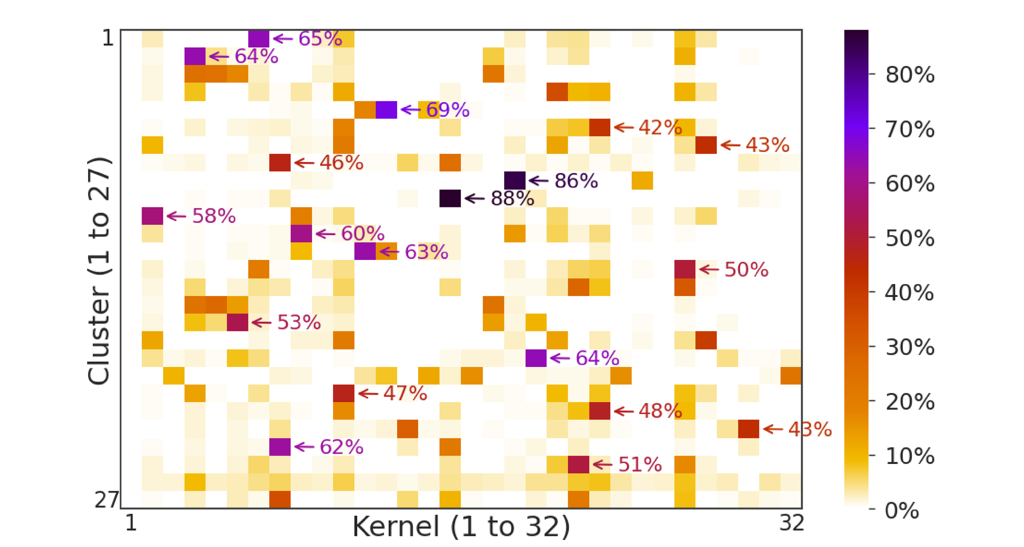

Figure 12. **Cluster sparsity.** Kernel activation distribution at 20 dB SNR across clusters. Contributions >40% are labeled. Mean cluster Gini = 0.8745.