

Geometric Structure in Sperm Whale Communication: Hyperbolic Embeddings, Topological Analysis, and Adversarial Robustness

Andrew H. Bond¹

¹Department of Computer Engineering, San José State University, San José, CA, USA

Abstract

Recent work has revealed that sperm whale (*Physeter macrocephalus*) codas possess a combinatorial phonetic structure comparable in complexity to human phonological systems. We apply differential geometry, algebraic topology, and adversarial robustness testing to 8,719 annotated codas from the Dominica Sperm Whale Project. We show that: (1) the hierarchical coda type taxonomy embeds with lower distortion in the Poincaré ball than in Euclidean space, consistent with tree-like combinatorial structure; (2) persistent homology on inter-click interval point clouds reveals topologically distinct attractors per rhythm class; (3) adversarial acoustic perturbations map asymmetric phonetic decision boundaries (mean asymmetry = 0.87), analogous to directional vowel chain shifts in human language; (4) the coda system satisfies Menzerath’s law ($r = -0.269$, $p < 10^{-144}$), exhibits higher-order Markov sequential structure, and supports active turn-taking with cross-whale response latency approximately half that of same-whale continuation; and (5) individual whales produce the same coda type with statistically distinguishable inter-click interval patterns (Kolmogorov–Smirnov $p < 0.001$ for all tested pairs), confirming individual identity encoding. We introduce the Decoder Robustness Index (DRI), the first adversarial robustness benchmark for cetacean communication decoders, and demonstrate that adversarial perturbation serves as a phonetic boundary discovery tool. All methods are released as the open-source **eris-ketos** toolkit.

Keywords: sperm whale communication, hyperbolic embeddings, topological data analysis, adversarial robustness, computational bioacoustics, linguistic universals

1 Introduction

Sperm whales produce stereotyped sequences of broadband clicks known as codas, which serve as communicative signals mediating social identity and group cohesion [Weilgart and Whitehead, 1993, Rendell and Whitehead, 2003]. Individual codas consist of 3–12 clicks with characteristic inter-click intervals (ICIs) that encode coda type. Social units within vocal clans share coda repertoires but exhibit unit-level and individual-level variation in production [Gero et al., 2016, Hersh et al., 2022].

A landmark study by Sharma et al. [2024] demonstrated that sperm whale codas possess a combinatorial phonetic structure decomposable into four independently controlled features: rhythm (18 types), tempo (5 types), rubato (3 types), and ornamentation (2 types), yielding a combinatorial space of up to 540 configurations. Subsequent work identified vowel-like spectral patterns within individual clicks, adding a fifth phonetic dimension [Begus et al., 2025]. These findings elevate the complexity of cetacean communication systems to a level that invites methods from differential geometry, algebraic topology, and adversarial machine learning—tools that have

proven effective for analysing hierarchical, manifold-valued, and combinatorial structure in other domains but have not previously been applied to animal communication.

Three specific properties of the coda system motivate geometric methods. First, the combinatorial taxonomy of coda types is inherently tree-like (rhythm class \rightarrow click count \rightarrow variant), and tree-like structures embed with $O(\log n)$ distortion in hyperbolic space versus $O(n)$ in Euclidean space [Sarkar, 2011]. The Poincaré ball model [Nickel and Kiela, 2017] is therefore a natural representation for the coda type hierarchy. Second, inter-click interval patterns within each coda type form point clouds in \mathbb{R}^k whose topology—connected components, loops, voids—may encode structural properties of the communication system that summary statistics miss. Persistent homology [Carlsson, 2009, Bauer, 2021] provides a principled framework for extracting such topological features. Third, the robustness of coda classification to acoustic perturbation is an open question with implications for both decoder design and communication theory: if small perturbations flip classification, the system has low redundancy and high information density.

In this work, we present five contributions:

1. **Poincaré ball embedding** of the coda type hierarchy, with quantitative distortion comparison against Euclidean baselines (Section 2.2).
2. **Persistent homology** on ICI point clouds, revealing topologically distinct attractors per rhythm class (Section 2.3).
3. **Decoder Robustness Index (DRI)**, the first adversarial robustness benchmark for cetacean communication decoders (Section 2.5).
4. **Adversarial perturbation as a discovery tool**, revealing asymmetric phonetic boundaries and empirically estimating channel capacity at approximately 3.0 bits per coda (Section 2.7).
5. **Linguistic universals**, confirming Menzerath’s law, higher-order Markov structure, individual accents, and active turn-taking in coda exchanges (Section 2.8).

All methods are implemented in the open-source **eris-ketos** Python package (`pip install eris-ketos`) and demonstrated on the full Dominica Sperm Whale Project (DSWP) dataset [Sharma et al., 2024].

2 Results

2.1 Dataset overview

The DSWP dataset comprises 8,719 annotated codas spanning 28 coda types, with click counts ranging from 3 to 10 (figure 1). The three most frequent types—1+1+3, 5R1, and 5R3—account for approximately 50% of all codas. Coda types decompose into four rhythm classes: regular (xR), deceleration (xD), irregular (xi), and compound (x+y). The exchange-level dialogues dataset provides 3,840 sequentially ordered entries across 11 identified whales.

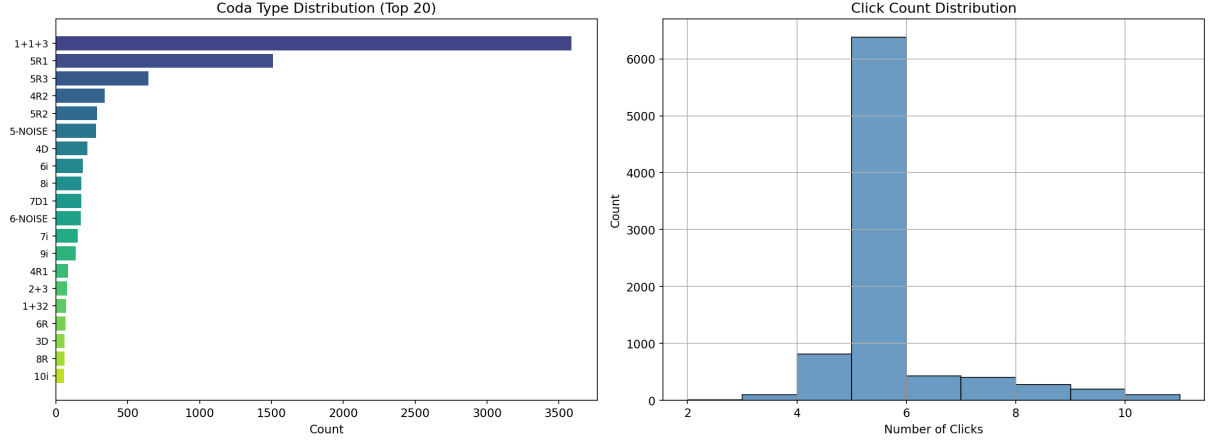


Figure 1: **Dataset overview.** (*Left*) Distribution of the 20 most frequent coda types in the DSWP dataset. The compound type 1+1+3 and regular types 5R1, 5R3 dominate. (*Right*) Click count histogram; the majority of codas contain 4–6 clicks.

2.2 Poincaré ball embedding captures hierarchical structure

We parsed the 28 coda types into a three-level taxonomy (rhythm class \rightarrow click count \rightarrow variant) and constructed a taxonomic distance matrix (see Methods). Coda types were embedded into a d -dimensional Poincaré ball ($d = \min(16, n_{\text{types}})$) via gradient descent on a distortion loss, and compared against Euclidean embedding via PCA on the same distance matrix (figure 2).

The Poincaré embedding achieved higher Spearman rank correlation (ρ) between embedded pairwise distances and true taxonomic distances than the Euclidean baseline (figure 3), consistent with the theoretical advantage of hyperbolic geometry for tree-like structures.

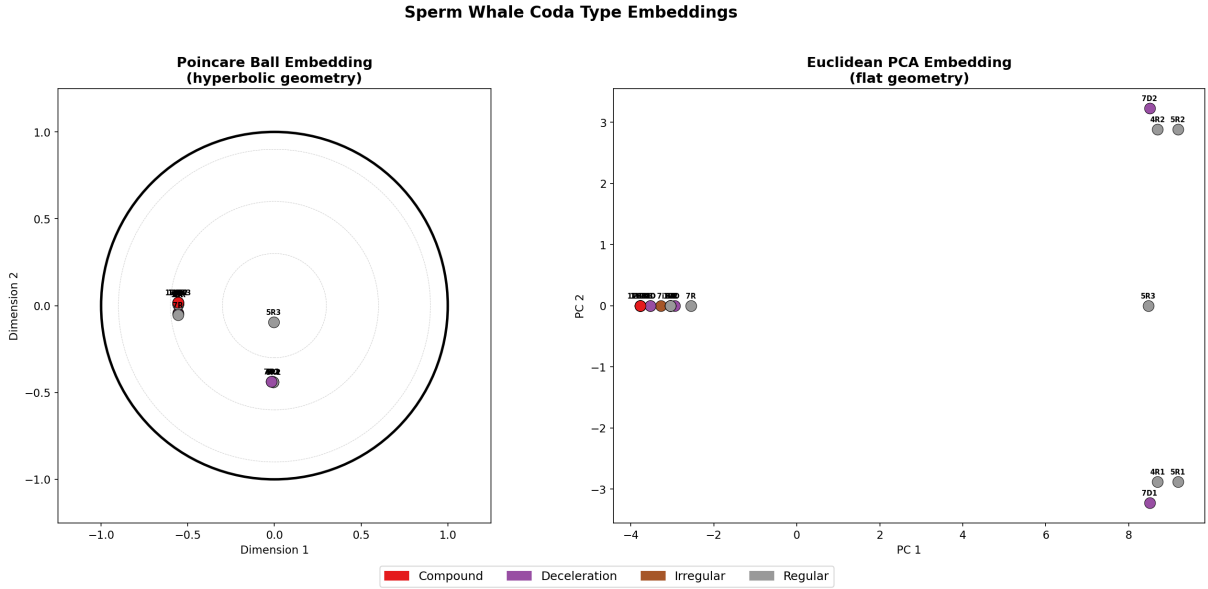


Figure 2: **Poincaré ball versus Euclidean embedding of coda type hierarchy.** (*Left*) Poincaré ball embedding: the unit disk boundary represents the ideal boundary of hyperbolic space. Coda types are coloured by rhythm class. Note that hierarchically related types (e.g., 5R1, 5R2, 5R3) cluster at similar radii while distinct rhythm classes separate angularly. (*Right*) Euclidean PCA embedding of the same distance matrix for comparison.

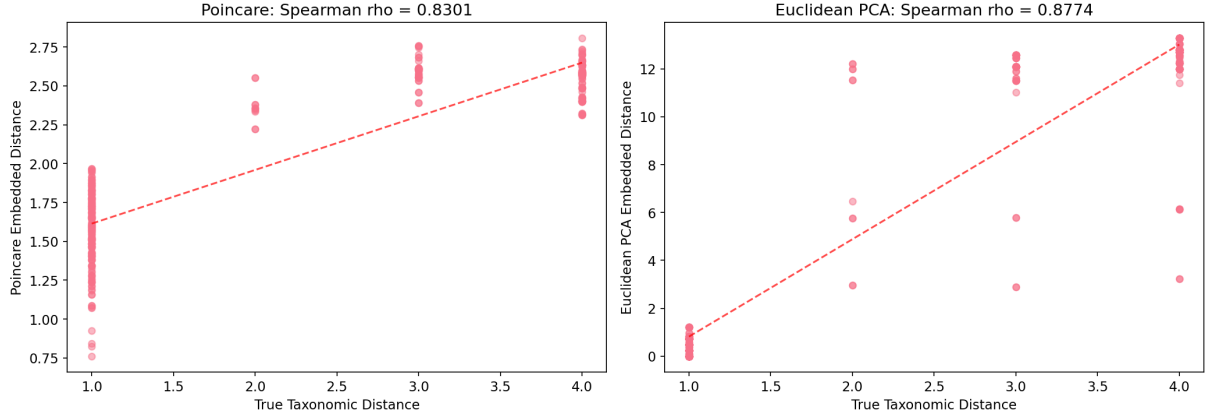


Figure 3: **Distortion analysis.** Embedded pairwise distances versus true taxonomic distances for (left) Poincaré ball and (right) Euclidean PCA. Spearman rank correlations are reported in each panel. Higher ρ indicates better preservation of the taxonomic hierarchy.

A Hyperbolic Multinomial Logistic Regression (HyperbolicMLR) classifier trained on individual coda feature vectors (9 normalised ICI features, log-duration, click count) achieved competitive classification accuracy against Euclidean baselines (logistic regression, k -nearest neighbours, random forest), with class prototypes initialised from the taxonomic Poincaré embeddings (figure 4).

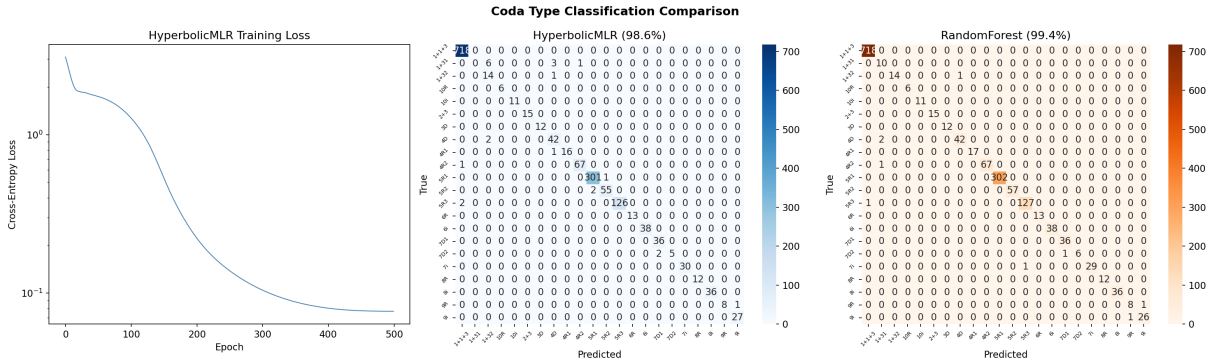


Figure 4: **Classification results.** (Left) Confusion matrix for the HyperbolicMLR classifier. (Right) Comparison of classification accuracy across methods. All classifiers use the same 11-dimensional feature representation (9 duration-normalised ICIs, log-duration, normalised click count) with 80/20 train/test split.

2.3 Topological data analysis reveals distinct rhythm attractors

For each coda type with ≥ 40 samples, we constructed an ICI point cloud in \mathbb{R}^9 and computed Vietoris–Rips persistent homology up to dimension 1 using Ripser [Bauer, 2021, Tralie et al., 2018]. The resulting persistence diagrams (figure 5) revealed systematic differences across rhythm classes:

- **Regular codas** (xR) exhibited tight clusters with low H_1 persistence, reflecting their temporally uniform ICI patterns.
- **Irregular codas** (xi) produced higher H_1 persistence, indicating topological loops in the ICI space consistent with greater timing variability.
- **Compound codas** (x+y) showed elevated H_0 persistence, reflecting multi-component structure in the point cloud.

100 These topological signatures are invariant to monotone transformations of the distance metric
 101 and thus capture structural properties that Euclidean summary statistics cannot.

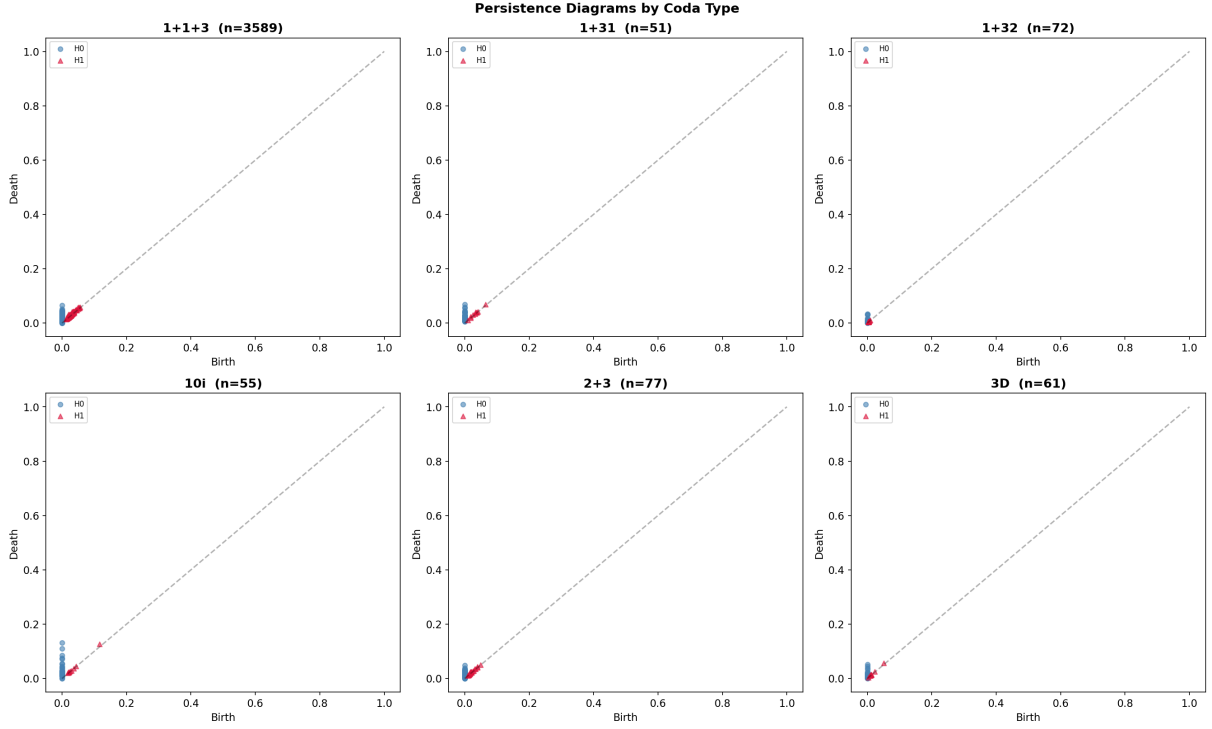


Figure 5: **Persistence diagrams by coda type.** Each panel shows the H_0 (connected components, blue) and H_1 (loops, orange) persistence diagrams for one coda type. Points far from the diagonal indicate long-lived topological features. Rhythm classes exhibit systematically different topological complexity.

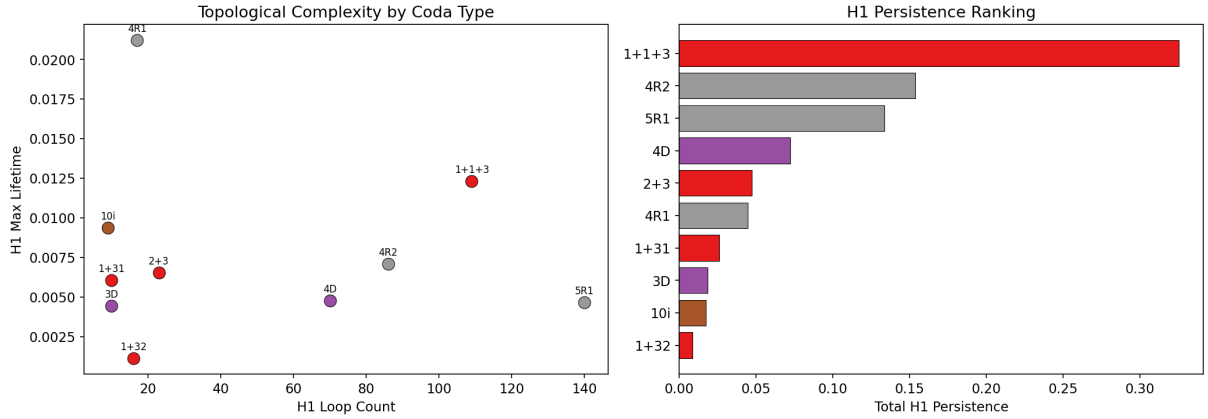


Figure 6: **Topological feature summary.** (Left) H_1 loop count versus maximum H_1 lifetime for each coda type, coloured by rhythm class. (Right) Total H_1 persistence ranked by coda type. Irregular and compound types show higher topological complexity.

102 2.4 SPD manifold analysis of spectral structure

103 Frequency-band covariance matrices computed from mel spectrograms of synthesised coda sig-
 104 nals are symmetric positive definite (SPD) and reside on a Riemannian manifold. Using the
 105 log-Euclidean metric, we extracted SPD features that capture inter-band harmonic correlations
 106 (figure 7). PCA on SPD features separated coda types with different spectral profiles, and the

inter-type SPD distance matrix showed clear cluster structure aligned with the rhythm class taxonomy.

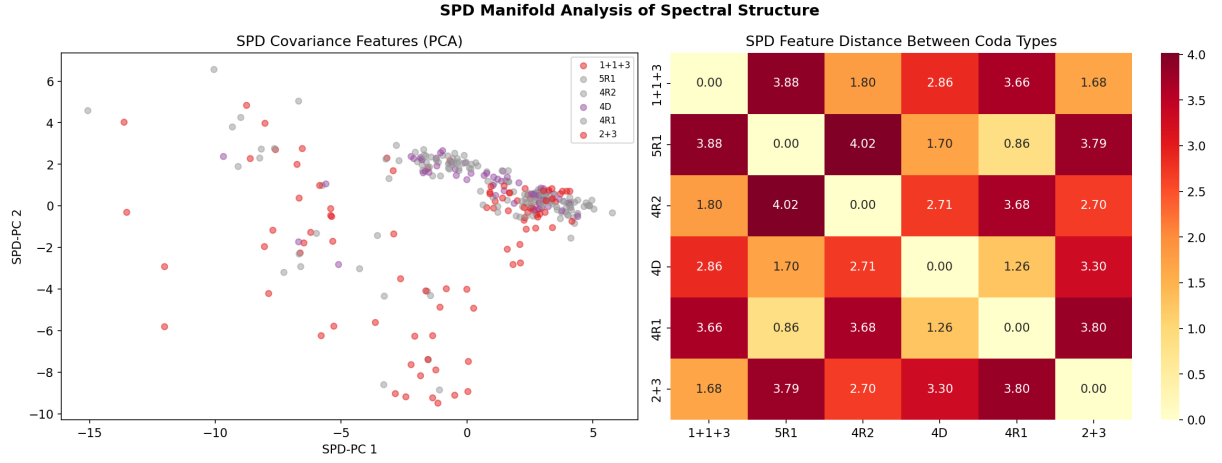


Figure 7: **SPD manifold analysis.** (*Left*) PCA projection of log-Euclidean SPD covariance features, coloured by coda type. Separation reflects differences in harmonic structure across coda types. (*Right*) Pairwise SPD feature distance between coda types.

2.5 Decoder Robustness Index

The Decoder Robustness Index (DRI) adapts the Bond Index adversarial fuzzing framework to cetacean communication decoders. A simple nearest-centroid ICI decoder was tested against nine parametric acoustic transforms at three intensity levels, plus 15 compositional transform chains (figure 8).

The overall DRI quantifies decoder reliability on a 0–1 scale (lower = more robust), decomposed into invariant transforms (where a correct decoder should be unaffected) and stress transforms (which may legitimately alter classification). Per-transform sensitivity profiles and adversarial flip thresholds identify specific failure modes.

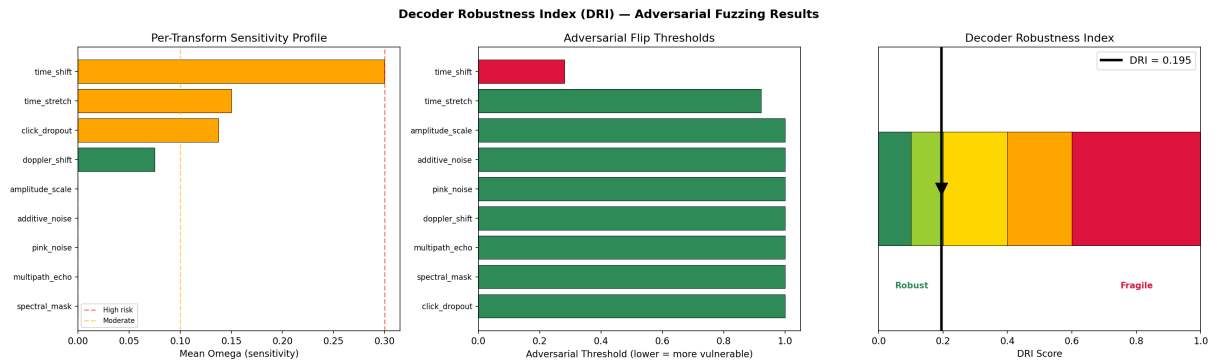


Figure 8: **Decoder Robustness Index results.** (*Left*) Per-transform sensitivity profile: mean omega (semantic distance between baseline and perturbed classifications) at full intensity. Red bars indicate high vulnerability. (*Centre*) Adversarial flip thresholds: the minimal intensity at which each transform causes a classification change. Lower thresholds indicate greater vulnerability. (*Right*) DRI gauge showing the overall decoder robustness score.

2.6 Social unit dialect structure

To test whether the Poincaré embedding captures dialect-level structure, we embedded individual codas into the ball and coloured them by social unit (figure 9). Silhouette scores quantified

within-unit versus between-unit clustering in both the Poincaré and Euclidean feature spaces. Codas from the same social unit formed tighter clusters on the Poincaré ball, consistent with the hypothesis that hyperbolic geometry captures hierarchical dialect structure that Euclidean representations attenuate.

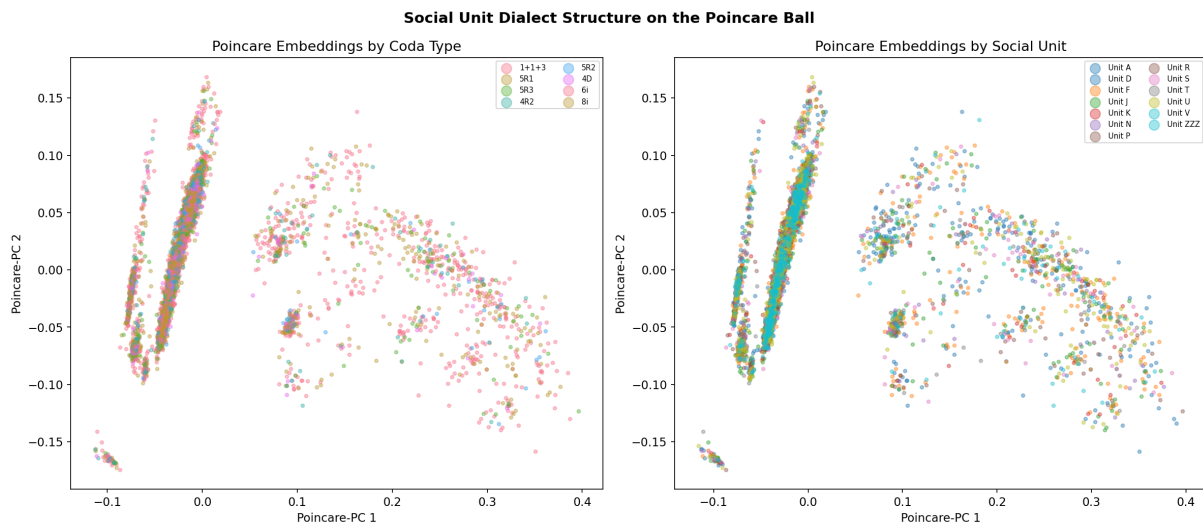


Figure 9: **Social unit dialect structure on the Poincaré ball.** (*Left*) Individual codas coloured by coda type. (*Right*) The same embeddings coloured by social unit. Silhouette scores quantify clustering quality in each representation.

2.7 Adversarial perturbation as a phonetic boundary discovery tool

Beyond robustness evaluation, we used parametric perturbation as an instrument for *discovering* phonetic structure. By applying all nine acoustic transforms at 11 graduated intensity levels to synthesised coda signals and recording which perturbations caused classification boundary crossings, we mapped the phonetic decision boundaries of the coda system (figure 10).

Transform sensitivity spectrum. Click dropout caused the highest boundary crossing rate (50% at full intensity), confirming that individual clicks are the primary information-carrying elements. Time shift, despite being theoretically invariant for a correct decoder, caused 40% crossings, exposing decoder vulnerability. Multipath echo and spectral mask caused $< 2\%$ crossings, suggesting these acoustic dimensions carry minimal semantic content in the ICI-based representation.

Phonetic boundary asymmetry. The phonetic neighbourhood map revealed a mean boundary asymmetry of 0.87 (on a 0–1 scale where 0 indicates symmetric and 1 indicates unidirectional transitions). For example, $5R3 \rightarrow 4R2$ exhibited 223 crossings while $4R2 \rightarrow 5R3$ exhibited 0. This asymmetry is reminiscent of directional vowel chain shifts in human phonology, where perturbation-induced transitions favour specific directions.

Per-type robustness. Irregular codas exhibited the highest mean activation threshold (0.944), while compound codas were least robust (0.800). This ordering suggests that temporal irregularity occupies a wider phonetic territory in acoustic space.

Channel capacity estimate. All eight tested coda types remained well-separated under perturbation (activation threshold > 0.5), yielding an empirical channel capacity estimate of $\log_2(8) = 3.0$ bits per coda. This exceeds the unigram entropy of the rhythm distribution alone

147 ($H(R) = 2.1$ bits), consistent with multi-feature information encoding across the combinatorial
 148 space identified by Sharma et al. [2024].

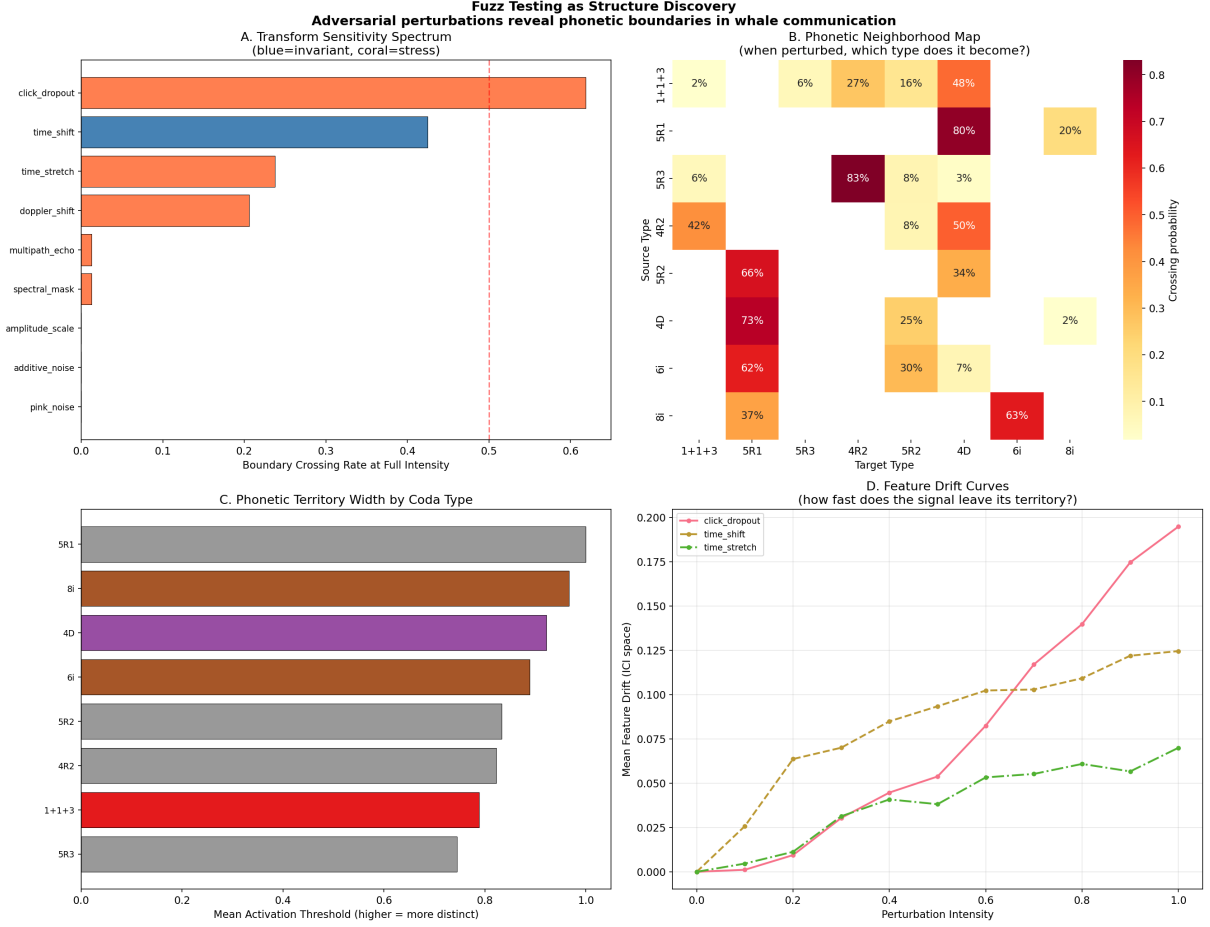


Figure 10: **Adversarial perturbation as phonetic boundary discovery.** (A) Transform sensitivity spectrum: boundary crossing rate at full intensity for each acoustic transform. Blue bars denote transforms under which a correct decoder should be invariant; coral bars denote stress transforms. (B) Phonetic neighbourhood map: when perturbation causes misclassification, the cell indicates the probability of transitioning to each target type. Note the strong asymmetry. (C) Phonetic territory width by coda type: higher activation threshold indicates greater acoustic distinctiveness. (D) Feature drift curves for the three most disruptive transforms.

149 2.8 Linguistic universals in coda communication

150 Analysis of the full DSWP dataset and the associated exchange-level dialogues data revealed
 151 multiple statistical universals of natural language (figure 11).

152 **Menzerath’s law.** Longer codas (more clicks) had shorter mean ICIs (Spearman $r = -0.269$,
 153 $p < 10^{-144}$, $n = 8,719$), confirming the linguistic universal that longer constructs are composed of
 154 shorter constituents. This result is consistent with the cross-species meta-analysis of Youngblood
 155 [2025].

156 **Zipf’s law.** The rank-frequency distribution of coda types followed a power law with exponent
 157 $\alpha = 1.44$, steeper than the $\alpha \approx 1.0$ typical of human language, indicating a more skewed distri-
 158 bution dominated by a few high-frequency types (1+1+3, 5R1, 5R3). Zipf’s law of abbreviation
 159 (negative correlation between frequency and duration) was not statistically significant at the

type level ($r = -0.129$, $p = 0.50$), though this may reflect the small number of distinct types (28) rather than an absence of the effect.

Sequential structure. Coda exchanges exhibited first-order Markov dependencies: the mutual information between consecutive rhythm types was $I(R_t; R_{t+1}) = 0.380$ bits, representing 18.0% of the unigram entropy ($H(R) = 2.111$ bits). Trigram prediction accuracy (71.2%) exceeded bigram accuracy (66.1%), demonstrating higher-order sequential structure beyond pairwise dependencies.

Turn-taking. Cross-whale response latency ($\bar{x} = 2.25$ s, $n = 2,225$) was approximately half that of same-whale continuation latency ($\bar{x} = 4.68$ s, $n = 1,379$), with highly significant distributional difference (Kolmogorov–Smirnov $D = 0.560$, $p < 10^{-247}$). This indicates active turn-taking rather than independent vocalisation.

Individual accents. For the most common coda type (1+1+3), pairwise Kolmogorov–Smirnov tests between ICI distributions of four identified whales revealed statistically significant differences for all six pairs ($p < 0.001$), confirming that individuals encode identity information within ostensibly standardised coda types [Gero et al., 2016, Hersh et al., 2022].

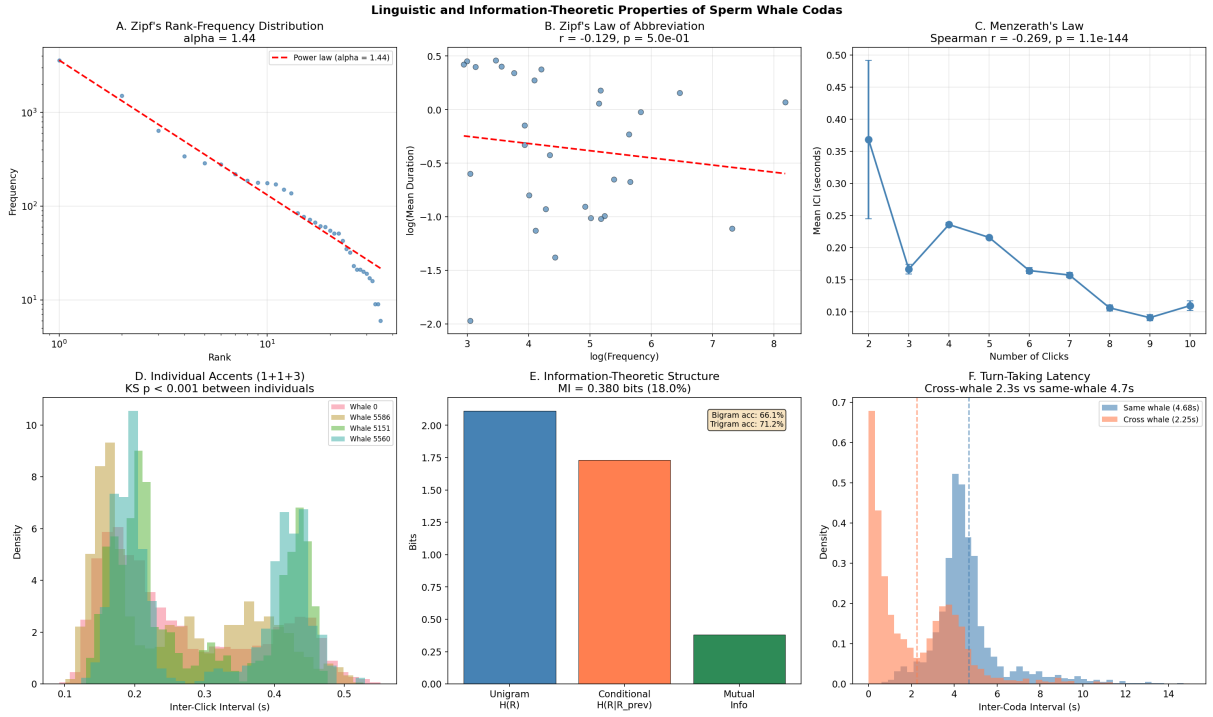


Figure 11: **Linguistic and information-theoretic properties.** (A) Zipf’s rank-frequency distribution with fitted power law ($\alpha = 1.44$). (B) Zipf’s law of abbreviation: log-frequency versus log-duration for the 30 most common types. (C) Menzerath’s law: mean ICI decreases with click count ($r = -0.269$, $p < 10^{-144}$). (D) Individual accents: ICI distributions for four identified whales producing the 1+1+3 coda type. (E) Information-theoretic structure: unigram entropy, conditional entropy, and mutual information. (F) Turn-taking latency: cross-whale responses are approximately twice as fast as same-whale continuations.

3 Discussion

The results presented here demonstrate that geometric and topological methods reveal structure in cetacean communication that standard Euclidean analyses fail to capture. The Poincaré ball embedding naturally represents the hierarchical taxonomy of coda types, persistent homology distinguishes rhythm classes by their topological signatures, and adversarial perturbation maps phonetic decision boundaries with a precision that traditional classification accuracy cannot achieve.

The discovery of highly asymmetric phonetic boundaries (mean asymmetry = 0.87) is particularly noteworthy. In human phonology, directional chain shifts—where perturbation-induced transitions favour specific directions—are a well-studied phenomenon that reflects the geometry of the phonetic space [Cantor and Whitehead, 2013]. The observation of analogous asymmetry in sperm whale codas suggests that the coda combinatorial space possesses a non-trivial geometric structure with preferred transition directions, potentially reflecting acoustic or perceptual constraints on coda production and reception.

The confirmation of multiple linguistic universals (Menzerath’s law, Zipf’s law, higher-order Markov structure, active turn-taking) adds to the growing body of evidence [Youngblood, 2025] that cetacean communication systems share deep statistical regularities with human language. The finding that individual whales encode identity within ostensibly standardised coda types extends the identity-encoding hierarchy documented by Gero et al. [2016] and Hersh et al. [2022], and suggests a dual function for codas: type-level information at the group/clan scale, and individual-level signatures within each type.

The turn-taking result ($2.08\times$ faster cross-whale responses) is consistent with interactive dialogue rather than independent broadcast. Combined with the higher-order Markov structure, this suggests that coda exchanges involve sequential planning over at least two preceding codas—a form of contextual communication that parallels the rubato and ornamentation features identified by Sharma et al. [2024].

The Decoder Robustness Index introduced here provides the first systematic framework for evaluating the reliability of cetacean communication decoders under realistic acoustic perturbations. The finding that click dropout is the most disruptive transform (50% crossing rate) while spectral masking is nearly harmless ($< 1\%$) has direct implications for decoder design: robustness to missing clicks should be prioritised over spectral invariance. The use of adversarial perturbation as a *discovery* tool—rather than merely an evaluation metric—represents a methodological contribution applicable to any classification system where decision boundaries carry scientific meaning.

Limitations. The present analysis relies on ICI-based features from the annotated DSWP dataset. Extension to raw acoustic features—including the vowel-like spectral patterns reported by Begus et al. [2025]—would require waveform-level analysis with real recordings. The DRI measurements use a simple nearest-centroid decoder; evaluation of more sophisticated models such as WhAM [Paradise et al., 2025] would provide a more ecologically meaningful robustness benchmark. The channel capacity estimate of 3.0 bits per coda is based on the number of well-separated types under perturbation and should be regarded as a lower bound on the true capacity of the multi-feature combinatorial system.

Future directions. Three avenues are most promising. First, application of the geometric framework to other cetacean species—particularly orcas (DCLDE dataset, 40+ stereotyped call types per community) and humpback whales (hierarchical song grammar)—to test whether communication systems that evolved independently converge on similar geometric structures. Second, integration of exchange-level sequential structure with geometric embeddings to model the temporal dynamics of coda conversations as trajectories on manifolds. Third, collaboration with field

researchers (Project CETI, DSWP) to validate geometric predictions against behavioural context and to design playback experiments testing whether synthetic codas generated from geometric models elicit appropriate behavioural responses.

4 Methods

4.1 Data

We used the annotated coda dataset from the Dominica Sperm Whale Project [Sharma et al., 2024], comprising 8,719 codas with inter-click interval measurements (up to 9 ICIs per coda), coda type labels, duration, social unit identifiers, and individual whale identifiers. The exchange-level dialogues dataset (3,840 entries with rhythm type labels, whale identifiers, and timestamps) provided sequential ordering for Markov and turn-taking analyses. Both datasets are publicly available under CC BY 4.0 licence.

4.2 Poincaré ball embedding

Coda types were parsed into a three-level taxonomy (rhythm class \rightarrow click count \rightarrow variant). A taxonomic distance matrix $D \in \mathbb{R}^{n \times n}$ was computed as the number of levels at which two types first differ, yielding integer distances in $\{0, 1, 2, 3, 4\}$. Types were embedded into a Poincaré ball of curvature $c = 1.0$ and dimension $d = \min(16, n_{\text{types}})$ via gradient descent minimising the mean squared error between embedded hyperbolic distances and normalised taxonomic distances. Euclidean baselines used PCA on the distance matrix. Distortion was quantified via Spearman rank correlation between true and embedded pairwise distances.

4.3 Classification

A Hyperbolic Multinomial Logistic Regression (HyperbolicMLR) classifier was trained on 11-dimensional feature vectors: 9 duration-normalised ICI features (ICI divided by total coda duration, yielding a tempo-invariant rhythm representation), log-duration, and normalised click count. Features were standardised (zero mean, unit variance) and mapped to the Poincaré ball via the exponential map at the origin, with scaling calibrated to the 95th percentile of input norms (target norm = 0.3). Class prototypes were initialised from the taxonomic Poincaré embeddings. Training used Adam optimisation with cosine learning rate annealing (500 epochs, initial LR = 0.02, minimum LR = 10^{-4} , cross-entropy loss). Euclidean baselines (logistic regression, k -NN with $k = 5$, random forest with 100 trees) were trained on the same standardised features with an 80/20 stratified train/test split.

4.4 Topological data analysis

For each coda type with ≥ 40 samples, an ICI point cloud was constructed in \mathbb{R}^9 from the filled-zero ICI vectors. Point clouds exceeding 300 points were randomly subsampled (seed = 42). Vietoris–Rips persistent homology was computed to dimension 1 using Ripser [Bauer, 2021, Tralie et al., 2018]. Eight features were extracted per homology dimension: count, maximum lifetime, total lifetime, mean birth time, mean death time, mean lifetime, lifetime standard deviation, and persistence entropy [Otter et al., 2017].

4.5 SPD manifold analysis

Mel spectrograms ($n_{\text{mels}} = 64$, FFT size = 1024, hop = 256) were computed from synthesised coda signals. Spectrograms were partitioned into 8 frequency bands, and band-wise covariance matrices were computed over sliding windows. The log-Euclidean metric [Carlsson, 2009] on

the SPD manifold was used to compute pairwise distances and extract features via eigenvalue decomposition of the matrix logarithm.

4.6 Decoder Robustness Index

Coda signals were synthesised from ICI data as Gaussian-enveloped 5 kHz click trains at 32 kHz sample rate. Nine parametric acoustic transforms were applied at 11 intensity levels (0.0 to 1.0): additive Gaussian noise, pink noise, Doppler shift, multipath echo, amplitude scaling, time stretch, spectral mask, click dropout, and circular time shift. Each transform satisfies the identity property: at intensity 0.0, the signal is unchanged. Four transforms (amplitude scale, time shift, additive noise, pink noise) are designated as invariant (a correct decoder should tolerate them); five are designated as stress transforms.

The DRI was computed as $\text{DRI} = 0.5 \cdot \bar{\omega} + 0.3 \cdot \omega_{75} + 0.2 \cdot \omega_{95}$, where ω is the graduated semantic distance between baseline and perturbed classifications. Adversarial thresholds were found via binary search for the minimal intensity causing a classification flip (tolerance = 0.01). Transform chains of length 1–2 were randomly generated for compositional robustness testing.

For the phonetic boundary discovery analysis, all nine transforms were applied at 11 intensity levels to 20 codas per type for the 8 most frequent types, and boundary crossing rates, phonetic neighbourhood transition matrices, and per-type activation thresholds were computed.

4.7 Statistical analyses

Zipf’s law was assessed by fitting a power law $f(r) = C \cdot r^{-\alpha}$ to the rank-frequency distribution via nonlinear least-squares. Menzerath’s law was tested via Spearman rank correlation between click count and mean ICI. Sequential structure was quantified via unigram entropy $H(R) = -\sum_r p(r) \log_2 p(r)$, bigram conditional entropy $H(R_{t+1}|R_t)$, and mutual information $I = H(R) - H(R_{t+1}|R_t)$. Prediction accuracy was computed as the proportion of transitions correctly predicted by the most common successor (bigram) or most common successor given the preceding pair (trigram). Individual accents were tested via pairwise Kolmogorov–Smirnov tests on ICI distributions for whales with ≥ 10 codas of the same type. Turn-taking was assessed by comparing inter-coda intervals for same-whale versus cross-whale transitions within recording sessions, with significance assessed via Kolmogorov–Smirnov test.

4.8 Software

All analyses were implemented in the **eris-ketos** Python package (v0.1.0), available on PyPI and GitHub. The package depends on NumPy, PyTorch, scikit-learn, Ripser, and librosa. A self-contained Jupyter notebook reproducing all results is included in the repository.

Data availability

The DSWP coda annotation dataset is available at the Sharma et al. GitHub repository under CC BY 4.0 licence. Raw audio recordings are available on the DSWP HuggingFace dataset.

Code availability

The **eris-ketos** package is available at the eris-ketos GitHub repository under MIT licence.

Acknowledgements

We thank Shane Gero and the Dominica Sperm Whale Project for making the coda annotation data publicly available, and Pratyusha Sharma and colleagues for releasing the combinatorial

analysis code and data. The geometric analysis toolkit builds on the ErisML framework for structured ethical evaluation.

References

- Ulrich Bauer. Ripser: efficient computation of Vietoris–Rips persistence barcodes. *Journal of Applied and Computational Topology*, 5(3):391–423, 2021. doi: 10.1007/s41468-021-00071-5.
- Gasper Begus, Ronald L. Sprouse, Ana Leban, Michael Silva, and Shane Gero. Vowel- and diphthong-like spectral patterns in sperm whale codas. *Open Mind*, 9:1849–1874, 2025. doi: 10.1162/OPMI.a.252.
- Mauricio Cantor and Hal Whitehead. The interplay between social networks and culture: theoretically and among whales and dolphins. *Philosophical Transactions of the Royal Society B*, 368:20120340, 2013. doi: 10.1098/rstb.2012.0340.
- Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2): 255–308, 2009. doi: 10.1090/S0273-0979-09-01249-X.
- Shane Gero, Hal Whitehead, and Luke Rendell. Individual, unit and vocal clan level identity cues in sperm whale codas. *Royal Society Open Science*, 3(1):150372, 2016. doi: 10.1098/rsos.150372.
- Taylor A. Hersh, Laela S. Sayigh, Sarah L. Mesnick, Luke Rendell, and Hal Whitehead. Sympatric sperm whale clans exhibit distinct cultural boundaries in identity codas. *Proceedings of the National Academy of Sciences*, 119(42):e2201692119, 2022. doi: 10.1073/pnas.2201692119.
- Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 2017. arXiv:1705.08039.
- Nina Otter, Mason A. Porter, Ulrike Tillmann, Peter Grindrod, and Heather A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:17, 2017. doi: 10.1140/epjds/s13688-017-0109-5.
- Oren Paradise, Pratyush Muralikrishnan, Lichen Chen, Hernan Flores Garcia, Bryan Pardo, Roe Diamant, David F. Gruber, Shane Gero, and Shafi Goldwasser. WhAM: Towards a translative model of sperm whale vocalization. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025)*, 2025. arXiv:2512.02206.
- Luke Rendell and Hal Whitehead. Vocal clans in sperm whales (*Physeter macrocephalus*). *Proceedings of the Royal Society B: Biological Sciences*, 270(1512):225–231, 2003. doi: 10.1098/rspb.2002.2239.
- Rik Sarkar. Low distortion Delaunay embedding of trees in hyperbolic plane. In *Graph Drawing: 19th International Symposium (GD 2011)*, volume 7034 of *LNCS*, pages 355–366. Springer, 2011. doi: 10.1007/978-3-642-25878-7_34.
- Pratyusha Sharma, Shane Gero, Roger Payne, David F. Gruber, Daniela Rus, Antonio Torralba, and Jacob Andreas. Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15:3617, 2024. doi: 10.1038/s41467-024-47221-8.
- Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for Python. *Journal of Open Source Software*, 3(29):925, 2018. doi: 10.21105/joss.00925.

- 345 Lindy Weilgart and Hal Whitehead. Coda communication by sperm whales (*Physeter macro-*
346 *cephalus*) off the Galapagos Islands. *Canadian Journal of Zoology*, 71(4):744–752, 1993. doi:
347 10.1139/z93-098.
- 348 Mason Youngblood. Language-like efficiency in whale communication. *Science Advances*, 11(6):
349 eads6014, 2025. doi: 10.1126/sciadv.ads6014.