Quorum-Theoretic Semantic Decoding Model

(QTSD)

A Comprehensive Framework for
Multi-Agent Acoustic Communication Analysis

SifSystems MCRD

December 2025

Abstract

This white paper presents the Quorum-Theoretic Semantic Decoding Model (QTSD), a rigorous mathematical framework for extracting semantic content from multi-agent acoustic communication systems. The framework treats cetacean communication as analogous to bacterial quorum sensing: accumulated acoustic density triggers collective behavioral transitions when crossing learned thresholds. Key innovations include factorized density computation achieving linear scaling, culture-agnostic embeddings enabling cross-population transfer, adversarial deconfounding for environmental robustness, and streaming architecture for real-world deployment. Three generative model candidates—Threshold-Crossing HMM, Neural Density Field Model, and Multi-Agent Communication Channel—are unified under a common data model and evaluation protocol, including causal validation through playback experiments.

Version 1.0 · Produced via Sif AI Orchestration System

Contents

1 Introduction and Motivation **2**
- 1.1 Core Hypothesis . . . 2
- 1.2 Framework Goals . . . 2

2 Symbol Table and Formal Ontology **2**
- 2.1 Core Symbol Definitions . . . 2
- 2.2 State Interpretation . . . 3

3 Data Model Specification **3**
- 3.1 Observed Data Structure . . . 3
- 3.2 Temporal Structure . . . 3
- 3.3 Hierarchical Structure . . . 3

4 Latent Structure Specification **3**
- 4.1 Conditional Dependencies . . . 3
- 4.2 Factorized Density Computation . . . 4

5 Acoustic Embedding Architecture **4**
- 5.1 Factorized Embedding Structure . . . 4
- 5.2 Training Objectives . . . 4

6 Generative Model Candidates **4**
- 6.1 Candidate A: Threshold-Crossing HMM (TC-HMM) . . . 4
- 6.2 Candidate B: Neural Density Field Model (NDFM) . . . 5
- 6.3 Candidate C: Multi-Agent Communication Channel (MACC) . . . 5

7 Inference Objectives **5**
- 7.1 Primary Objectives . . . 5
- 7.2 Composite Loss Function . . . 5

8 Identifiability Constraints **5**
- 8.1 Sources of Degeneracy . . . 5
- 8.2 Formal Constraints . . . 6

9 Experimental Protocols **6**
- 9.1 Protocol 1: Behavior Prediction from Density . . . 6
- 9.2 Protocol 2: Threshold Stability Across Groups . . . 6
- 9.3 Protocol 3: Playback Intervention (Causal Test) . . . 7
- 9.4 Protocol 4: Cross-Environment Generalization . . . 7
- 9.5 Protocol 5: Cross-Population Transfer . . . 7

10 Configuration Reference **7**
- 10.1 Core Hyperparameters . . . 7
- 10.2 Regularization Weights . . . 8

11 Discussion **8**
- 11.1 Strengths of the Framework . . . 8
- 11.2 Limitations and Future Directions . . . 8
- 11.3 Conclusion . . . 8

1 Introduction and Motivation

The central challenge in cetacean communication research is decoding semantic content from complex, naturalistic acoustic data. Traditional approaches search for discrete lexical units—"words" or "phrases"—but this may fundamentally mischaracterize how meaning emerges in these systems.

The QTSD framework proposes an alternative: communication as **accumulated acoustic density** that triggers collective behavior when crossing thresholds. This is directly analogous to quorum sensing in bacterial colonies, where accumulated autoinducer molecules trigger coordinated gene expression.

1.1 Core Hypothesis

Rather than individual whales transmitting discrete messages, many individuals vocalize, creating an acoustic "pressure" that builds until the group collectively shifts behavior. The semantic content lies not in individual calls but in the density field they create.

1.2 Framework Goals

1. **Scalability**: Linear complexity in population size and recording duration
2. **Robustness**: Invariance to environmental and recording conditions
3. **Generalization**: Transfer across populations with distinct cultural patterns
4. **Hierarchy**: Principled handling of Population $\supset$ Pod $\supset$ Matriline $\supset$ Individual structure
5. **Causality**: Validation protocols that establish causal, not merely correlational, relationships

2 Symbol Table and Formal Ontology

2.1 Core Symbol Definitions

Symbol	Description	Type
X_t^i	Raw acoustic features for individual i at time t	$\mathbb{R}^d$, Observable
ρ_t	Acoustic density field at time t	$\mathbb{R}^+$, Derived
$\phi(X)$	Call embedding function	$\mathbb{R}^d \to \mathbb{R}^k$, Learned
$K(\Delta t, \Delta x)$	Spatiotemporal influence kernel	$\mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^+$
S_t^i	Latent arousal/attention state	$\{0, 1, 2\}$, Latent
τ_b	Threshold for behavior type b	$\mathbb{R}^+$, Latent
B_t	Observable group behavior at time t	Categorical
C_t	Environmental context	$\mathbb{R}^m$, Observable
A_i	Individual identity	Categorical
Θ	Global parameters (species-level)	Parameter Space
N_t	Count of individuals in active state	$\mathbb{Z}^+$, Derived
e_t	Environment embedding	$\mathbb{R}^{d_{env}}$, Derived

2.2 State Interpretation

The latent state $S_t^i \in \{0, 1, 2\}$ represents:

- $S_t^i = 0$: **Monitoring** — Individual is listening, not vocalizing
- $S_t^i = 1$: **Active** — Individual is currently vocalizing
- $S_t^i = 2$: **Refractory** — Post-vocalization cooldown period

3 Data Model Specification

3.1 Observed Data Structure

The observed data takes the form:

$$D = \left\{(X_t^{1:N}, \mathbf{x}_t^{1:N}, B_t, C_t)\right\}_{t=1}^{T} \tag{1}$$

Where:

- $X_t^{1:N}$: Acoustic observations from N individuals (sparse; most entries null)
- $\mathbf{x}_t^{1:N}$: Spatial positions of individuals
- $B_t \in \{b_1, \ldots, b_K, \emptyset\}$: Group behavior label or null
- C_t: Context vector encoding environmental conditions

3.2 Temporal Structure

- **Indexing**: $t \in \mathbb{Z}^+$ (discrete, event-driven sampling)
- **Stochastic assumption**: Semi-Markov process with explicitly modeled dwell times
- **Duration bounds**: Maximum duration $D_{\max}$ (recommended: 64 frames)

3.3 Hierarchical Structure

Each level inherits parameters from its parent with low-rank deviations:

$$\theta_L = \theta_{\text{shared}} + \sum_{l=1}^{L} \Delta\theta_l \tag{2}$$

where $\Delta\theta_l = U_l \cdot v_l$ with $U_l \in \mathbb{R}^{D \times r_l}$ and rank $r_l \ll D$.

Recommended ranks: $r_{\text{population}} = 8$, $r_{\text{pod}} = 4$, $r_{\text{matriline}} = 2$, $r_{\text{individual}} = 1$.

4 Latent Structure Specification

4.1 Conditional Dependencies

The core probabilistic structure:

$$p(S_t^i \mid S_{t-1}^i, \rho_t, C_t, \Theta) \quad \text{(Individual state transitions)} \tag{3}$$

$$p(B_t \mid N_t, \rho_t, \{\tau_b\}, C_t) \quad \text{(Behavior emission)} \tag{4}$$

$$\rho_t = f(X_{1:t}^{1:N}, \mathbf{x}_{1:t}^{1:N}, \Theta_\phi, \Theta_K) \quad \text{(Density computation)} \tag{5}$$

4.2 Factorized Density Computation

The density computation employs a sliding window with exponential moving average:

$$\rho_t^i = \sum_{j \in \text{Neighbors}(i,t)} \sum_{\tau \in \text{Window}(t)} K(\phi(X_t^i), \phi(X_\tau^j)) \cdot w_\tau \tag{6}$$

Where:

- $\text{Window}(t) = [t - k_{\text{temporal}}, t]$ with $k_{\text{temporal}} \in [64, 256]$
- $\text{Neighbors}(i,t) = \text{TopK}_{\text{spatial}}$ with $k_{\text{spatial}} \in [8, 32]$
- $\text{EMA}_t = \alpha \cdot \text{EMA}_{t-1} + (1-\alpha) \cdot h_t$ with $\alpha \approx 0.99$

Complexity reduction: $O((NT)^2) \to O(N \cdot k_{\text{spatial}} \cdot k_{\text{temporal}} \cdot T)$

5 Acoustic Embedding Architecture

5.1 Factorized Embedding Structure

The embedding function decomposes into three components:

$$\phi(X) = \phi_{\text{universal}}(X) + \phi_{\text{cultural}}(X, \text{pop}_{id}) + \phi_{\text{env}}(X, e_t) \tag{7}$$

Universal Component (∼75% of dimensions):

- Fundamental frequency (F0) contours and derivatives
- Spectral centroid, bandwidth, and flux trajectories
- Amplitude envelope and modulation spectrum
- Harmonic-to-noise ratio and formant-like spectral peaks

Cultural Component (∼25% of dimensions):

$$\phi_{\text{cultural}}(X, \text{pop}_{id}) = U_{\text{pop}} \cdot v_{\text{pop}_{id}} \tag{8}$$

Environmental Component (Conditional Instance Normalization):

$$\phi_{\text{normalized}}(X) = \gamma(e_t) \cdot \frac{\phi_{\text{raw}}(X) - \mu(e_t)}{\sigma(e_t)} + \beta(e_t) \tag{9}$$

5.2 Training Objectives

$$L_{\text{embed}} = L_{\text{task}} + \lambda_{\text{pop\_adv}} \cdot L_{\text{pop\_adv}} + \lambda_{\text{cross\_pop}} \cdot L_{\text{cross\_pop}} + \lambda_{\text{primitive}} \cdot L_{\text{primitive}} \tag{10}$$

6 Generative Model Candidates

6.1 Candidate A: Threshold-Crossing HMM (TC-HMM)

$$p(D, S, \Theta) = p(\Theta) \prod_t \left[p(B_t \mid N_t, \rho_t, \tau) \prod_i p(X_t^i \mid S_t^i)\, p(S_t^i \mid S_{t-1}^i, \rho_t) \right] \tag{11}$$

Transition probability:

$$p(S_t^i = 1 \mid S_{t-1}^i = 0, \rho_t) = \sigma\left(\frac{\rho_t - \tau}{\epsilon}\right) \tag{12}$$

Behavior emission:

$$p(B_t = b \mid N_t, \rho_t) = \text{Categorical}\left(\text{softmax}(W_b \cdot [\rho_t, N_t])\right) \tag{13}$$

Type: Hidden Semi-Markov Model with density-dependent transitions.

6.2 Candidate B: Neural Density Field Model (NDFM)

$$p(D, Z, \Theta) = p(\Theta) \prod_t p(X_t \mid Z_t, \Theta_x)\, p(B_t \mid Z_t, \Theta_b)\, p(Z_t \mid Z_{t-1}, \rho_t, C_t, \Theta_z) \tag{14}$$

With GRU dynamics:

$$Z_t = \mathrm{GRU}(Z_{t-1}, [\rho_t, C_t]; \Theta_z) \tag{15}$$

Type: RNN with attention-based density computation, continuous latent space $Z_t \in \mathbb{R}^k$.

6.3 Candidate C: Multi-Agent Communication Channel (MACC)

$$p(D, S, M, \Theta) = p(\Theta) \prod_t \left[p(B_t \mid M_t) \prod_i p(X_t^i \mid S_t^i, M_t)\, p(S_t^i \mid S_{t-1}^i, M_{t-1}) \right] \prod_t p(M_t \mid M_{t-1}, \{S_t^i\}) \tag{16}$$

Shared message pool dynamics:

$$M_t = (1 - \gamma) M_{t-1} + \sum_i 1[S_t^i = \text{active}] \cdot \phi(X_t^i) \tag{17}$$

Type: Communication channel with explicit encoder/decoder structure.

7 Inference Objectives

7.1 Primary Objectives

(OBJ-1) Posterior Inference:

$$\hat{S}_{1:T}, \hat{Z}_{1:T} = \arg\max_{S,Z} p(S_{1:T}, Z_{1:T} \mid X_{1:T}, B_{1:T}, C_{1:T}, \hat{\Theta}) \tag{18}$$

(OBJ-2) Predictive Likelihood Maximization:

$$\Theta^* = \arg\max_{\Theta} \sum_{\text{sessions}} \log p_\Theta(B_{1:T} \mid X_{1:T}, C_{1:T}) \tag{19}$$

(OBJ-3) Density-Behavior Mutual Information:

$$\Theta^* = \arg\max_{\Theta} I_\Theta(\rho_{1:T}; B_{1:T} \mid C_{1:T}) \tag{20}$$

Subject to rate constraint: $I_\Theta(X_t; \phi(X_t)) \leq R$

7.2 Composite Loss Function

$$L(\Theta) = L_{\text{likelihood}} + \lambda_{\text{MI}} L_{\text{MI}} + \lambda_{\text{ARD}} L_{\text{ARD}} + \lambda_{\text{behavior}} L_{\text{behavior}} + \lambda_{\text{env\_adv}} L_{\text{env\_adv}} \tag{21}$$

8 Identifiability Constraints

8.1 Sources of Degeneracy

1. **Scale ambiguity**: $\phi' = c \cdot \phi$, $\tau' = c \cdot \tau$ yield equivalent models
2. **Individual-embedding confound**: Variation in τ_i versus ϕ_i
3. **Environmental confounds**: C_t may directly cause B_t
4. **Cultural bias**: Population-specific patterns in universal representations

8.2 Formal Constraints

(C1) Scale Normalization:

$$\mathbb{E}[\|\phi(X)\|_2^2] = 1 \tag{22}$$

(C2) Threshold Separation:

$$|\tau_b - \tau_{b'}| \geq \delta \quad \forall b \neq b' \tag{23}$$

(C3) Information Bottleneck:

$$I(\phi(X); X) \leq R, \quad I(\phi(X); B) \geq \beta \tag{24}$$

(C4) Context-Invariance:

$$\mathrm{Var}_C[\mathbb{E}[B \mid \rho, C]] \leq \eta \cdot \mathrm{Var}[\mathbb{E}[B \mid \rho]] \tag{25}$$

(C5) Temporal Consistency (Lipschitz):

$$\|Z_t - Z_{t+1}\|_2 \leq L \cdot \|\rho_t - \rho_{t+1}\|_2 \tag{26}$$

(C6) Threshold Sparsity:

$$|\{b : \tau_b \text{ is distinct}\}| \leq K_{\max} \tag{27}$$

(C7) Population Invariance:

$$I(\phi_{\text{universal}}(X); \text{pop}_{id}) \leq \epsilon_{\text{cultural}} \tag{28}$$

(C8) Environment Invariance:

$$I(Z_t; \text{env}_{id}) \leq \epsilon_{\text{env}} \tag{29}$$

9 Experimental Protocols

9.1 Protocol 1: Behavior Prediction from Density

Component	Specification
Train	$D_{\text{train}} = \{(\rho_t, B_t, C_t)\}$ from sessions $1, \ldots, n$
Test	D_{test} from held-out sessions
Metrics	$M_1 = \text{Accuracy}(\hat{B}_t, B_t)$, $M_2 = \text{AUC-ROC}$
Baseline	M_1^0: predict B_t from C_t alone
Success	$M_1 - M_1^0 \geq \delta_1$

9.2 Protocol 2: Threshold Stability Across Groups

Component	Specification
Train	Estimate $\hat{\tau}_b$ on pods $P_1, \ldots, P_k$
Test	Apply $\hat{\tau}_b$ to new pod P_{k+1}
Metric	$M_3 = \mathrm{Var}_{\text{pods}}[\hat{\tau}_b]/\mathbb{E}[\hat{\tau}_b]$ (CV)
Success	$M_3 \leq \epsilon_{\text{species}}$

9.3 Protocol 3: Playback Intervention (Causal Test)

Component	Specification
Intervention	Inject synthetic calls $\tilde{X}$ to raise ρ_t
Prediction	Model predicts $P(B_t = b \mid \rho_t + \Delta\rho)$
Observation	Measure actual B_t^{obs}
Metric	$M_4 = \text{KL}(P_{\text{predicted}} \| P_{\text{observed}})$
Success	$M_4 \leq \epsilon_{\text{causal}}$

9.4 Protocol 4: Cross-Environment Generalization

Component	Specification
Train	Subset of environments E_{train}
Test	Held-out environments E_{test}
Metric	$M_5 =$ Performance ratio $E_{\text{test}}/E_{\text{train}}$
Success	$M_5 \geq 0.85$ without TTA, ≥ 0.95 with TTA

9.5 Protocol 5: Cross-Population Transfer

Component	Specification
Train	Populations $\text{Pop}_1, \ldots, \text{Pop}_k$
Test	Novel population using $\phi_{\text{universal}}$ only
Metric	$M_6 =$ Zero-shot behavior prediction accuracy
Success	$M_6 \geq 0.7\times$ within-population performance

10 Configuration Reference

10.1 Core Hyperparameters

Parameter	Recommended	Description
k_{temporal}	64–256	Temporal window size
k_{spatial}	8–32	Maximum spatial neighbors
$D_{\max}$	32–128	Maximum HSMM duration
d_{embed}	128–512	Total embedding dimension
$d_{\text{universal}}$	$0.75 \times d_{\text{embed}}$	Universal embedding dimension
K_{neg}	16–256	Negatives per MI anchor
K_{beh}	16–64	Behavior prototypes
chunk_length	1024–8192	Training chunk size

10.2 Regularization Weights

Parameter	Default	Description
λ_{MI}	0.1	MI objective weight
λ_{ARD}	0.01	ARD prior weight
$\lambda_{\text{behavior}}$	0.1	Behavior prototype weight
$\lambda_{\text{env\_adv}}$	0.3	Environment adversarial weight
$\lambda_{\text{pop\_adv}}$	0.5	Population adversarial weight
$\lambda_{\text{ind\_adv}}$	0.3	Individual adversarial weight
β_{IB}	0.01	Information bottleneck weight

11 Discussion

11.1 Strengths of the Framework

1. The quorum-sensing analogy is biologically plausible and offers a departure from lexical approaches
2. Identifiability analysis anticipates and addresses mathematical degeneracies
3. Experimental protocols include causal validation, not just correlational measures
4. Multiple model candidates enable principled model selection
5. Streaming architecture addresses real-world deployment requirements

11.2 Limitations and Future Directions

1. **Call-type specificity**: Current formulation pools all vocalizations; call-type-specific density fields may be needed
2. **Directed communication**: No modeling of who responds to whom
3. **Silence as information**: Only vocalizations are modeled; silence patterns may carry semantic content
4. **Temporal microstructure**: Call-internal modulation patterns not captured
5. **Baseline comparisons**: Simpler models (e.g., call count) needed for rigorous evaluation

11.3 Conclusion

The QTSD framework provides a mathematically rigorous, implementation-ready specification for decoding cetacean communication. By treating vocalizations as contributions to a collective density field rather than discrete messages, it offers a fundamentally different approach to the semantic decoding problem. The inclusion of causal validation protocols—particularly playback experiments—provides a path to distinguishing genuine communicative function from mere correlation.

This document was produced through the Sif system in under 2 hours, demonstrating AI-accelerated R&D capabilities.