

WOOF IS A METAPING

A Discrete-Packet, Information-Theoretic Framework
for Computational Canine Bioacoustics

Prepared by the AutoSciencePro5 Pipeline ■ June 2026

Domain: Computational Bioacoustics & Information Theory ■ Status: Theoretical White Paper, Revision 2.0

Scope and Epistemic Status

This paper proposes an *analytical lens*, not a claim of literal physiological machinery. To keep the speculative content separable from the settled science, every claim below falls into one of three tiers:

- **Established (E)**. Empirically supported acoustics and ethology, with citations: the source-filter model of voice production; formant dispersion as a body-size cue mediated by vocal-tract length; and context/arousal information carried by canine vocalisations.
- **Formalism (F)**. Standard, uncontroversial signal-processing mathematics (windowing, harmonic synthesis, linear predictive coding) recast in a network/telemetry vocabulary. The mathematics is sound; the *vocabulary* is a deliberate metaphor.
- **Speculative (S)**. Conjectural mappings offered as hypotheses, explicitly flagged in the text. These are *not* established and are stated here so they can be tested or rejected (see §6).

Terminology note. “Quantised” is used throughout in the loose sense of *discrete and time-localised*, not in the quantum-mechanical sense; there is no quantum physics in this model.

Abstract

Classical ethology tends to treat the domestic dog’s bark (*Canis lupus familiaris*) as an analogue, graded emotional outburst. **(E)** A substantial body of work shows, however, that barks and related calls carry structured, decodable information—about the caller’s body size, individual identity, and behavioural context. This paper offers a unifying reading of those findings: we model a single bark as a discrete, time-bounded acoustic *packet*—a *Metaping*—and ask what “header fields” it carries and how. **(F)** Using a windowed source-filter decomposition, we show that one vocal event simultaneously encodes (i) a size-correlated formant signature, (ii) an arousal-correlated fundamental-frequency trajectory, and (iii) a harshness/valence term governed by the harmonic-to-noise ratio. **(S)** We then sketch—explicitly as conjecture—a “network” layer in which inter-bark timing acts as an adaptive transmission cadence and a received packet perturbs a listener’s behavioural state. The contribution is framing, not new data: the value is a precise, falsifiable vocabulary that connects established bioacoustics to modern machine-listening pipelines.

1. Introduction and Framing

A long-standing problem in mammalian bioacoustics is how listeners extract *structural* facts—roughly how large the caller is, and where it is—almost instantly and across very different acoustic environments. Purely affective models (mapping “fear,” “aggression,” “play” onto duration, pitch, and harshness) describe the emotional surface but do not, on their own, explain this fast extraction of physical attributes.

We adopt a complementary framing borrowed from communications engineering. Rather than treating the bark as an unstructured analogue stream, we treat each event as a *packetised transmission*: a self-delimiting unit that carries a compact metadata header. The analogy to an Internet Control Message Protocol (ICMP) echo (“ping”) is deliberate but limited. An ICMP

echo mainly verifies reachability; the *Metaping* additionally broadcasts a header whose physical fields—chiefly the formant pattern—are constrained by the sender’s anatomy and are therefore *difficult, though not impossible, to fake* (§3.2). The packet/header language is a lens **(F)**; the underlying size and context cues are real **(E)**.

2. The Discrete Wave Packet (Time-Domain Analysis)

A bark is distinguished from continuous calls (growls, whines) by a fast onset, a high peak amplitude, and a rapid decay. Because of those sharp temporal boundaries we model the bark not as a stationary signal but as a single, time-localised energy packet. Let $W(t)$ be the radiated pressure wave:

$$W(t) = \Pi(t) S(t), \quad (1)$$

where $S(t)$ is the internally generated core signal (§3) and $\Pi(t)$ is an asymmetric temporal gate enforcing the packet’s boundaries from onset t_{start} to resolution. We write the gate as a time-varying Gaussian envelope multiplied by a causal (Heaviside) switch $H(\cdot)$:

$$\Pi(t) = \exp\left(-\frac{(t - t_0)^2}{2\sigma^2(t)}\right) H(t - t_{\text{start}}), \quad \sigma(t) = \begin{cases} \sigma_a, & t \leq t_0 \quad (\text{sharp attack}), \\ \sigma_d > \sigma_a, & t > t_0 \quad (\text{slower decay}). \end{cases} \quad (2)$$

Here t_0 is the instant of peak amplitude; the inequality $\sigma_d > \sigma_a$ reproduces the characteristic attack–decay asymmetry of a bark. **(F)** This is ordinary signal windowing written in packet language.

3. The Metadata Manifest: Source–Filter Deconstruction

The core signal $S(t)$ is governed by the *source–filter* theory of voice production **(E)**, which separates the glottal energy source from the resonances of the upper vocal tract:

$$S(t) = (E * V)(t) + N(t), \quad (3)$$

where $*$ denotes convolution, $E(t)$ is the glottal source, V the vocal-tract filter, and $N(t)$ an additive turbulent-noise term. Each component carries a distinct “header field.”

3.1 Glottal source $E(t)$ — the arousal field

Vocal-fold vibration produces a quasi-periodic pulse train whose fundamental frequency f_0 defines the carrier:

$$E(t) = \sum_{k=1}^K A_k(t) \cos\left(2\pi k \int_0^t f_0(\tau) d\tau + \varphi_k\right). \quad (4)$$

The instantaneous trajectory of $f_0(t)$ tracks *arousal and motivational state* rather than body size **(E)**. This distinction matters and is a correction to a common confusion: across many species f_0 is *not* a reliable indicator of caller size, whereas formant spacing is (Fitch, 1997). Consistent with Morton’s motivation–structural rules, low and harsh phonation tends to accompany agonistic/assertive states while higher, rising f_0 tends to accompany alarm or appeasement. In packet terms, f_0 behaves like a priority/urgency flag, not a size field.

3.2 Vocal-tract filter $V(z)$ — the biometric field

As the glottal pulse train passes through the pharyngeal and oral cavities it is shaped by an acoustic filter whose resonant peaks are the *formants*. In the discrete Z -domain this is modelled, as standard, by an all-pole linear-predictive-coding (LPC) transfer function **(F)**:

$$V(z) = \frac{1}{1 - \sum_{i=1}^P a_i z^{-i}}. \quad (5)$$

The poles (roots of the denominator) give the formant frequencies $F_1, F_2, \dots, F_N$. Modelling the tract as a uniform tube of length L , the mean spacing between successive formants—the *formant dispersion* ΔF —is approximately constant and inversely proportional to L (**E**) (Fitch, 1997):

$$\Delta F = \frac{1}{N-1} \sum_{i=1}^{N-1} (F_{i+1} - F_i) \approx \frac{c}{2L}, \quad (6)$$

with $c \approx 350 \text{ m s}^{-1}$ the speed of sound in warm air. Because vocal-tract length scales with skeletal size, ΔF supports an allometric estimate of body mass (**E**) (Riede & Fitch, 1999; Taylor, Reby & McComb, 2008):

$$M = \alpha \left(\frac{c}{2\Delta F} \right)^\beta = \alpha L^\beta, \quad (7)$$

where α, β are fitted constants. *Honesty of the field*. Because ΔF is set largely by fixed anatomy, the size field is hard to counterfeit, which is why anatomical constraints are said to “generate honesty” in such signals (Reby & McComb, 2003). It is, however, *not* strictly unforgeable: dogs and other mammals can partly lower formants by extending the vocal tract (e.g. retracting the larynx), producing modest, documented size exaggeration. The correct claim is therefore *generally honest and costly to fake*, not *unforgeable*.

3.3 Noise term $N(t)$ — the harshness/valence field

To signal urgency or threat, phonation can be driven into a chaotic, less-periodic regime. We model the noise as $N(t) = \gamma(t)\xi(t)$, with $\xi(t)$ zero-mean Gaussian noise and $\gamma(t)$ an envelope set by the harmonic-to-noise ratio HNR. A low HNR (harsh, noisy phonation) is associated with agonistic/defensive states; a high HNR (clear, tonal phonation) with calmer or affiliative states (**E**). This is a perceptual correlate of harshness, not a hard switch.

4. A Network Reading of Multi-Packet Behaviour

The fields above can be collected into a single *telemetry vector* Ψ_W summarising one packet. Table 1 maps each acoustic parameter to its proposed field and tier.

Table 1: Proposed Metaping field map. E = established, F = formalism, S = speculative.

Acoustic parameter	Proposed field	Interpretation	Tier
Fundamental freq. f_0	Arousal / urgency flag	Motivational state and its rate of change; <i>not</i> a size cue.	E
Formant dispersion ΔF	Biometric (size) field	Vocal-tract length $\rightarrow$ body size; costly to fake.	E
Harmonic-to-noise ratio HNR	Harshness / valence	Tonal (calm/affiliative) vs. noisy (agonistic).	E
Inter-bark interval (IBI)	Transmission cadence	Tracks context and arousal; bark <i>rate</i> carries context.	E/S

4.1 Temporal structure and the cadence field

Empirically, much of the *context* in dog vocal behaviour lives in temporal structure—bark rate, rhythm, and spacing—rather than in the spectrum of an isolated call (**E**) (Pongrácz et al., 2005; Taylor, Reby & McComb). We therefore read the inter-bark interval (IBI) as an adaptive transmission cadence: a node that is more aroused, or addressing a nearer target, tends to transmit more densely.

4.2 Speculative extension: echoic and receiver-state effects

The following two mappings are **Speculative (S)** and are stated as hypotheses, not findings.

(S1) *Echoic sensing.* For any emitted sound, the round-trip delay to a reflector at distance d is $\tau_e = 2d/c$, and reverberation does carry passive information about the environment. It is tempting to read IBI modulation as active, sonar-like mapping of local obstacles. We flag this as **unsupported**: dogs are *not* known to echolocate—active biosonar is restricted to odontocetes, several bats, some shrews, and a couple of cave-dwelling birds. The more parsimonious reading of IBI is arousal/context (§4.1), and the echoic reading should be treated as a falsifiable conjecture, not a mechanism.

(S2) *Receiver-state update.* As a toy model of how a packet influences a listener j at distance d_{ij} , let the listener’s behavioural-state vector X_j update by

$$X_j(t + \tau_p) = X_j(t) + \Omega \left[\frac{\Psi_W}{d_{ij}^2} \right], \quad \tau_p = \frac{d_{ij}}{c}, \quad (8)$$

where τ_p is the propagation delay and Ω a receiver-sensitivity matrix. The inverse-square term is at least physically motivated (acoustic intensity falls as $1/d^2$ under spherical spreading); the linear state-update itself is a modelling convenience, not a measured law.

5. Application: Machine Listening

Reading the bark as “anatomical structure + intentional payload” suggests a concrete engineering move: normalise out the size field before decoding intent. Adding an allometric normalisation layer based on Eq. (7) removes a large, body-size-driven source of between-individual variance, which should let a single shared model generalise across breeds.

This is consistent with a real and active line of work **(E)**: self-supervised speech encoders pre-trained on human audio—such as WavLM and HuBERT—transfer surprisingly well to canine vocalisations, improving dog recognition, breed and sex classification, and context grounding (Abzaliev, Pérez Espinosa & Mihalcea, 2024; Chen et al., 2022). One caveat on terminology: published results to date rely on *transfer learning and linear probing* / *fine-tuning* in low-resource settings, not strict *zero-shot* decoding; the latter remains aspirational.

6. Limitations and Falsifiable Predictions

The framework earns its keep only if it makes testable predictions. Four are immediate:

1. **Size field.** ΔF estimated via Eq. (6) should predict measured vocal-tract length and body mass within the error bars reported by Fitch (1997) and Riede & Fitch (1999). If it does not, Eq. (7) is mis-specified.
2. **Arousal vs. size separation.** Manipulating f_0 while holding ΔF fixed should shift perceived arousal but *not* perceived size; manipulating ΔF alone should do the reverse. (Playback designs of this type already exist for growls.)
3. **Cadence, not sonar (test of S1).** IBI should covary with behavioural context and arousal but should *not* systematically track obstacle distance d in a controlled room. A null result on the distance term *falsifies* the echoic-sensing conjecture while leaving the rest of the framework intact.
4. **Normalisation gain.** An allometric normalisation layer should measurably reduce cross-breed error on intent-classification tasks relative to an unnormalised baseline.

Conclusion

The bark is not literally a network packet, and nothing here is quantum. What the “Metaping” framing buys is *precision*: it forces a clean separation between an anatomically constrained size field, an arousal-driven fundamental, a harshness/valence term, and a temporal cadence—each tied, where possible, to established results, and each speculative extension marked as such and paired with a way to refute it. That separation is exactly the structure a modern machine-listening pipeline wants, which is the framework’s most defensible practical claim.

References

1. Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
2. Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, 102(2), 1213–1222. doi:10.1121/1.421048.
3. Riede, T., & Fitch, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *Journal of Experimental Biology*, 202(20), 2859–2867. doi:10.1242/jeb.202.20.2859.
4. Taylor, A. M., Reby, D., & McComb, K. (2008). Human listeners attend to size information in domestic dog growls. *Journal of the Acoustical Society of America*, 123(5), 2903–2909. doi:10.1121/1.2896962.
5. Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, 65(3), 519–530.
6. Pongrácz, P., Molnár, C., Miklósi, Á., & Csányi, V. (2005). Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *Journal of Comparative Psychology*, 119(2), 136–144. doi:10.1037/0735-7036.119.2.136.
7. Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855–869. doi:10.1086/283219.
8. Chen, S., Wang, C., Chen, Z., *et al.* (2022). WavLM: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, 16(6), 1505–1518. doi:10.1109/JSTSP.2022.3160049.
9. Abzaliev, A., Pérez Espinosa, H., & Mihalcea, R. (2024). Towards dog bark decoding: leveraging human speech processing for automated bark classification. *Proceedings of LREC-COLING 2024*. arXiv:2404.18739.