Emergent intensity invariance in a physiologically inspired model of the grasshopper auditory system

Jona Hartling, Jan Benda

1 Exploring a grashopper's sensory world

Strong dependence on acoustic signals for ranged communication
- Diverse species-specific sound repertoires and production mechanisms
- Different contexts/ranges: Stridulatory, mandibular, wings, walking sounds
- Mate attraction/evaluation, rival deterrence, loss-of-signal predator alarm
→ Elaborate acoustic behaviors co-depend on reliable auditory perception

Songs = Amplitude-modulated (AM) broad-band acoustic signals
- Generated by stridulatory movement of hindlegs against forewings
- Shorter time scales: Characteristic temporal waveform pattern
- Longer time scales: High degree of periodicity (pattern repetition)
- Sound propagation: Signal intensity varies strongly with distance to sender
- Ectothermy: Temporal structure warps with temperature
→ Sensory constraints imposed by properties of the acoustic signal itself

Multi-species, multi-individual communally inhabited environments
- Temporal overlap: Simultaneous singing across individuals/species common
- Frequency overlap: No/hardly any niche speciation into frequency bands
- "Biotic noise": Hetero-/conspecifics ("Another one's songs are my noise")
- "Abiotic noise": Wind, water, vegetation, anthropogenic
- Effects of habitat structure on sound propagation (landscape - soundscape)
→ Sensory constraints imposed by the (acoustic) environment

Cluster of auditory challenges (interlocking constraints → tight coupling):
From continuous acoustic input, generate neuronal representations that...
1)...allow for the separation of relevant (song) events from ambient noise floor

2)...compensate for behaviorally non-informative song variability (invariances)
3)...carry sufficient information to characterize different song patterns, recognize the ones produced by conspecifics, and make appropriate behavioral decisions based on context (sender identity, song type, mate/rival quality)

How can the auditory system of grasshoppers meet these challenges?
- What are the minimum functional processing steps required?
- Which known neuronal mechanisms can implement these steps?
- Which and how many stages along the auditory pathway contribute?

→ What are the limitations of the system as a whole?

How can a human observer conceive a grasshopper's auditory percepts?
- How to investigate the workings of the auditory pathway as a whole?
- How to systematically test effects and interactions of processing parameters?
- How to integrate the available knowledge on anatomy, physiology, ethology?

→ Abstract, simplify, formalize → Functional model framework

Precursor work for model construction (special thanks to authors):

Linear-nonlinear modelling of behavioral responses to artificial songs
- Feature expansion as implemented in our model: Major contribution!
- Bank of linear filters, nonlinearity, temporal integration, feature weighting

→ Clemens and Hennig 2013 (crickets)
→ Clemens and Ronacher 2013 (grasshoppers)
→ Ronacher et al. 2015

Own advancements/key differences:
1) Used boxcar functions as artificial "songs" (focus on few key parameters)
→ Now actual, variable songs (as naturalistic as possible)
2) Fitted filters to behavioral data
→ More general, simpler, unfitted formalized Gabor filter bank

2 Developing a functional model of the grasshopper auditory pathway

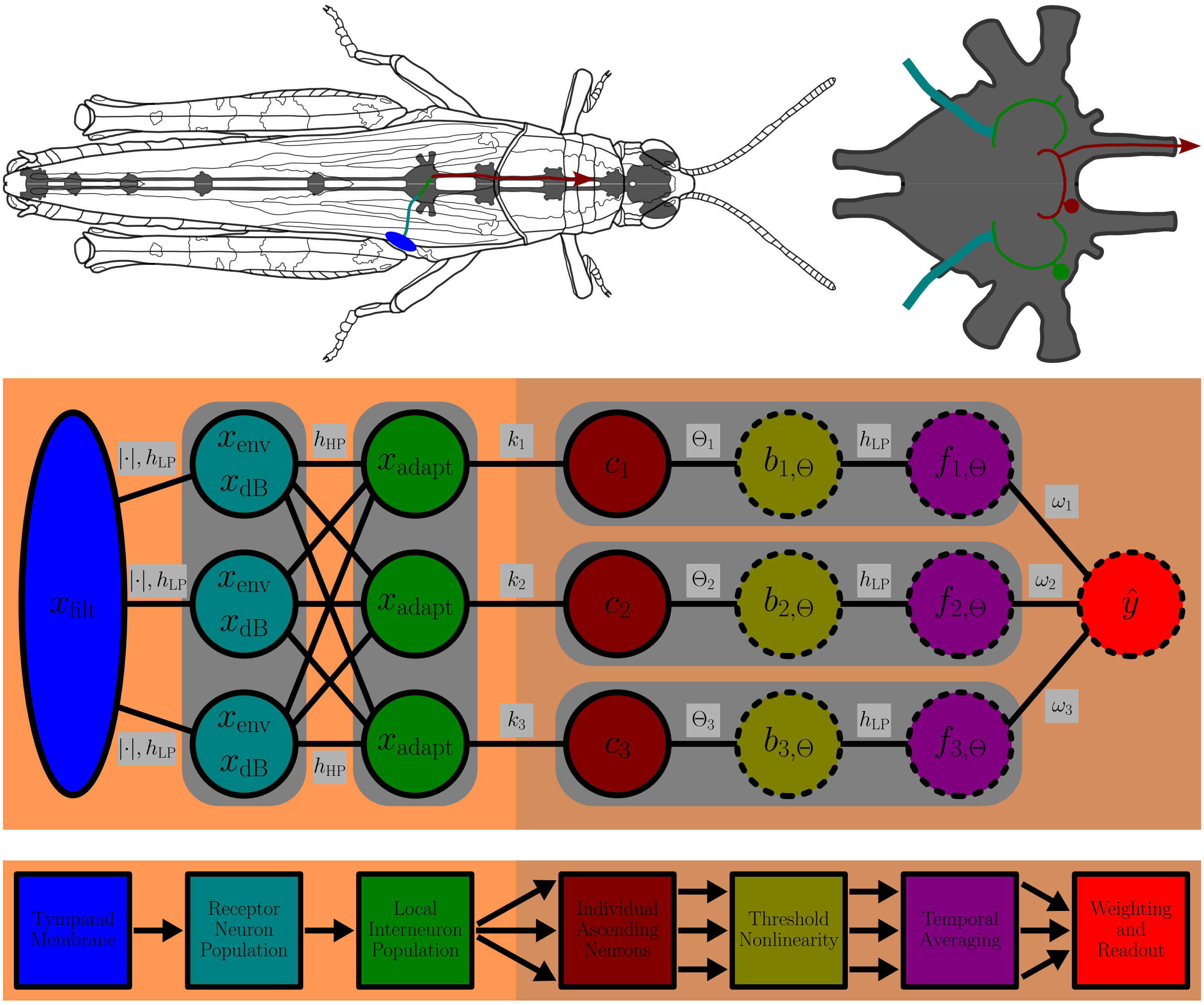

Figure 1: **The auditory system of grasshoppers.**

2.1 Population-driven signal pre-processing

Grasshoppers receive airborne sound waves by a tympanal organ at each side of the thorax (Fig. 1a). The tympanal membrane acts as a mechanical resonance filter, that focuses vibrations of specific frequencies on different membrane areas while attenuating others (Michelsen 1971; Windmill et al. 2008; Malkin et al. 2014). This processing step can be approximated by an initial bandpass filter

$$x_{\text{filt}}(t) = x(t) * h_{\text{BP}}(t), \qquad f_{\text{cut}} = 5\,\text{kHz},\ 30\,\text{kHz} \tag{1}$$

applied to the acoustic input signal $x(t)$. The auditory receptor neurons connect directly to the tympanal membrane. Besides performing the mechano-electrical transduction, the receptor population further is substrate to several known processing steps. First, the receptors extract the signal envelope (Machens, Prinz, et al. 2001), which likely involves a rectifying nonlinearity (Machens, Stemmler, et al. 2001). This can be modelled as full-wave rectification followed by lowpass filtering

$$x_{\mathrm{env}}(t) \;=\; |x_{\mathrm{filt}}(t)| \,*\, h_{\mathrm{LP}}(t), \qquad f_{\mathrm{cut}} \;=\; 500\,\mathrm{Hz} \tag{2}$$

of the tympanal signal $x_{\mathrm{filt}}(t)$. Furthermore, the receptors exhibit a sigmoidal response curve over logarithmically compressed intensity levels (Suga 1960; Gollisch et al. 2002). In the model, logarithmic compression is achieved by conversion to decibel scale

$$x_{\mathrm{dB}}(t) \;=\; 10 \cdot \log_{10} \frac{x_{\mathrm{env}}(t)}{x_{\mathrm{ref}}}, \qquad x_{\mathrm{ref}} \;=\; \max[x_{\mathrm{env}}(t)] \tag{3}$$

relative to the maximum intensity x_{ref} of the signal envelope $x_{\mathrm{env}}(t)$. The axons of the receptor neurons project into the metathoracic ganglion, where they synapse onto local interneurons (Fig. 1b). Both the local interneurons (Hildebrandt et al. 2009; Clemens, Weschke, et al. 201 and, to a lesser extent, the receptors themselves (Fisch et al. 2012) display spike-frequency adaptation in response to sustained stimulation. This behavior is crucial to render subsequent signal representations invariant to variations in sound intensity.

”Pre-split portion” of the auditory pathway:
Tympanal membrane → Receptor neurons → Local interneurons

Similar response/filter properties within receptor/interneuron populations (Clemens, Kutzki, et al. 2011)
→ One population-wide response trace per stage (no ”single-cell resolution”)

Stage-specific processing steps and functional approximations:

Initial: Continuous acoustic input signal $x(t)$

Filtering of behaviorally relevant frequencies by tympanal membrane
→ Bandpass filter 5-30 kHz

Extraction of signal envelope (AM encoding) by receptor population
→ Full-wave rectification, then lowpass filter 500 Hz

Logarithmically compressed intensity tuning curve of receptors
→ Decibel transformation

Spike-frequency adaptation in receptor and interneuron populations

→ Highpass filter 10 Hz

$$x_{\text{adapt}}(t) \,=\, x_{\text{dB}}(t) \,*\, h_{\text{HP}}(t), \qquad f_{\text{cut}} \,=\, 10\,\text{Hz} \tag{4}$$

2.2 Feature extraction by individual neurons

"Post-split portion" of the auditory pathway:
Ascending neurons (AN) → Central brain neurons

Diverse response/filter properties within AN population (Clemens, Kutzki, et al. 2011)
- Pathway splitting into several parallel branches
- Expansion into a decorrelated higher-dimensional sound representation

→ Individual neuron-specific response traces from this stage onwards

Stage-specific processing steps and functional approximations:

Template matching by individual ANs
- Filter base (STA approximations): Set of Gabor kernels
- Gabor parameters: σ_i, ϕ_i, f_i → Determines kernel sign and lobe number

$$k_i(t,\, \sigma_i,\, f_i,\, \phi_i) \,=\, e^{-\frac{t^2}{2\sigma_i{}^2}} \,\cdot\, \sin(2\pi f_i \,\cdot\, t \,+\, \phi_i) \tag{5}$$

→ Separate convolution with each member of the kernel set

$$c_i(t) \,=\, x_{\text{adapt}}(t) \,*\, k_i(t) = \int_{-\infty}^{+\infty} x_{\text{adapt}}(\tau) \,\cdot\, k_i(t \,-\, \tau)\, d\tau \tag{6}$$

Thresholding nonlinearity in ascending neurons (or further downstream)
- Binarization of AN response traces into "relevant" vs. "irrelevant"

→ Shifted Heaviside step-function $H(c_i \,-\, \Theta_i)$ (or steep sigmoid threshold?)

$$b_i(t,\, \Theta_i) \,=\, \begin{cases} 1, & c_i(t) \,>\, \Theta_i \\ 0, & c_i(t) \,\leq\, \Theta_i \end{cases} \tag{7}$$

Temporal averaging by neurons of the central brain
- Finalized set of slowly changing kernel-specific features (one per AN)
- Different species-specific song patterns are characterized by a distinct combination of feature values → Clusters in high-dimensional feature space

→ Lowpass filter 1 Hz

$$f_i(t) \,=\, b_i(t) \,*\, h_{\mathrm{LP}}(t), \qquad f_{\mathrm{cut}} \,=\, 1\,\mathrm{Hz} \tag{8}$$

3 Two mechanisms driving the emergence of intensity-invariant song representation

3.1 Logarithmic scaling & spike-frequency adaptation

Envelope $x_{\mathrm{env}}(t) \xrightarrow{\mathrm{dB}}$ Logarithmic $x_{\mathrm{dB}}(t) \xrightarrow{h_{\mathrm{HP}}(t)}$ Adapted $x_{\mathrm{adapt}}(t)$

- Rewrite signal envelope $x_{\mathrm{env}}(t)$ (Eq. 2) as a synthetic mixture:
1) Song signal $s(t)$ ($\sigma_{\mathrm{s}}^2 = 1$) with variable multiplicative scale $\alpha \geq 0$
2) Fixed-scale additive noise $\eta(t)$ ($\sigma_{\eta}^2 = 1$)

$$x_{\mathrm{env}}(t) \,=\, \alpha \,\cdot\, s(t) \,+\, \eta(t), \qquad x_{\mathrm{env}}(t) \,>\, 0 \;\forall\; t \,\in\, \mathbb{R} \tag{9}$$

- Signal-to-noise ratio (SNR): Ratio of variances of synthetic mixture $x_{\mathrm{env}}(t)$ with ($\alpha > 0$) and without ($\alpha = 0$) song signal $s(t)$, assuming $s(t) \perp \eta(t)$

$$\mathrm{SNR} \,=\, \frac{\sigma_{s+\eta}^2}{\sigma_{\eta}^2} \,=\, \frac{\alpha^2 \,\cdot\, \sigma_{\mathrm{s}}^2 \,+\, \sigma_{\eta}^2}{\sigma_{\eta}^2} \,=\, \alpha^2 \,+\, 1 \tag{10}$$

Logarithmic component:

- Simplify decibel transformation (Eq. 3) and apply to synthetic $x_{\mathrm{env}}(t)$
- Isolate scale α and reference x_{ref} using logarithm product/quotient laws

$$\begin{aligned} x_{\mathrm{dB}}(t) \,&=\, \log \frac{\alpha \,\cdot\, s(t) \,+\, \eta(t)}{x_{\mathrm{ref}}} \\ &=\, \log \frac{\alpha}{x_{\mathrm{ref}}} \,+\, \log b_i g[s(t) \,+\, \frac{\eta(t)}{\alpha} b_i g] \end{aligned} \tag{11}$$

→ In log-space, a multiplicative scaling factor becomes additive
→ Allows for the separation of song signal $s(t)$ and its scale α
→ Introduces scaling of noise term $\eta(t)$ by the inverse of α
→ Normalization by x_{ref} applies equally to all terms (no individual effects)

Adaptation component:

- Highpass filter over $x_{\mathrm{dB}}(t)$ (Eq. 4) can be approximated as subtraction of the local signal offset within a suitable time interval T_{HP} $(0 \ll T_{\text{HP}} < \frac{1}{f_{\text{cut}}})$

$$x_{\text{adapt}}(t) \;\approx\; x_{\text{dB}}(t) \;-\; \log\frac{\alpha}{x_{\text{ref}}} \;=\; \log b_i g[s(t) \;+\; \frac{\eta(t)}{\alpha} b_i g] \tag{12}$$

Implication for intensity invariance:

- Logarithmic scaling is essential for equalizing different song intensities
→ Intensity information can be manipulated more easily when in form of a signal offset in log-space than a multiplicative scale in linear space

- Scale α can only be redistributed, not entirely eliminated from $x_{\text{adapt}}(t)$
→ Turn initial scaling of song $s(t)$ by α into scaling of noise $\eta(t)$ by $\frac{1}{\alpha}$

- Capability to compensate for intensity variations, i.e. selective amplification of output $x_{\text{adapt}}(t)$ relative to input $x_{\text{env}}(t)$, is limited by input SNR (Eq. 10):
$\alpha \gg 1$: Attenuation of $\eta(t)$ term → $s(t)$ dominates $x_{\text{adapt}}(t)$
$\alpha \approx 1$ Negligible effect on $\eta(t)$ term → $x_{\text{adapt}}(t) = \log[s(t) + \eta(t)]$
$\alpha \ll 1$: Amplification of $\eta(t)$ term → $\eta(t)$ dominates $x_{\text{adapt}}(t)$
→ Ability to equalize between different sufficiently large scales of $s(t)$
→ Inability to recover $s(t)$ when initially masked by noise floor $\eta(t)$

- Logarithmic scaling emphasizes small amplitudes (song onsets, noise floor)
→ Recurring trade-off: Equalizing signal intensity vs preserving initial SNR

3.2 Threshold nonlinearity & temporal averaging

Convolved $c_i(t) \xrightarrow{H(c_i\,-\,\Theta_i)}$ Binary $b_i(t) \xrightarrow{h_{\text{LP}}(t)}$ Feature $f_i(t)$

Thresholding component:

- Within an observed time interval T, $c_i(t)$ follows probability density $p(c_i,\, T)$
- Within T, $c_i(t)$ exceeds threshold value Θ_i for time T_1 $(T_1 + T_0 = T)$
- Threshold $H(c_i \,-\, \Theta_i)$ splits $p(c_i,\, T)$ around Θ_i in two complementary parts

$$\int_{\Theta_i}^{+\infty} p(c_i, T)\, dc_i \;=\; 1 \;-\; \int_{-\infty}^{\Theta_i} p(c_i, T)\, dc_i \;=\; \frac{T_1}{T} \tag{13}$$

→ Semi-definite integral over right-sided portion of split $p(c_i,\, T)$ gives ratio of time T_1 where $c_i(t) > \Theta_i$ to total time T due to normalization of $p(c_i,\, T)$

$$\int_{-\infty}^{+\infty} p(c_i,\, T)\, dc_i \;=\; 1 \tag{14}$$

Averaging component:

- Lowpass filter over binary response $b_i(t)$ (Eq. 8) can be approximated as temporal averaging over a suitable time interval T_{LP} ($T_{\mathrm{LP}} > \frac{1}{f_{\mathrm{cut}}}$)
- Within T_{LP}, $b_i(t)$ takes a value of 1 ($c_i(t) > \Theta_i$) for time T_1 ($T_1 + T_0 = T_{\mathrm{LP}}$)

$$f_i(t) \approx \frac{1}{T_{\mathrm{LP}}} \int_t^{t + T_{\mathrm{LP}}} b_i(\tau)\, d\tau = \frac{T_1}{T_{\mathrm{LP}}} \tag{15}$$

→ Temporal averaging over $b_i(t) \in [0, 1]$ (Eq. 7) gives ratio of time T_1 where $c_i(t) > \Theta_i$ to total averaging interval T_{LP}
→ Feature $f_i(t)$ approximately represents supra-threshold fraction of T_{LP}

Combined result:

- Feature $f_i(t)$ can be linked to the distribution of $c_i(t)$ using Eqs. 13 & 15

$$f_i(t) \approx \int_{\Theta_i}^{+\infty} p(c_i, T_{\mathrm{LP}})\, dc_i = P(c_i > \Theta_i, T_{\mathrm{LP}}) \tag{16}$$

→ Because the integral over a probability density is a cumulative probability, the value of feature $f_i(t)$ (temporal compression of $b_i(t)$) at every time point t signifies the probability that convolution output $c_i(t)$ exceeds the threshold value Θ_i during the corresponding averaging interval T_{LP}

Implication for intensity invariance:

- Convolution output $c_i(t)$ quantifies temporal similarity between amplitudes of template waveform $k_i(t)$ and signal $x_{\mathrm{adapt}}(t)$ centered at time point t
→ Based on amplitudes on a graded scale

- Feature $f_i(t)$ quantifies the probability that amplitudes of $c_i(t)$ exceed threshold value Θ_i within interval T_{LP} around time point t
→ Based on binned amplitudes corresponding to one of two categorical states → Deliberate loss of precise amplitude information
→ Emphasis on temporal structure (ratio of T_1 over T_{LP})

- Thresholding of $c_i(t)$ and subsequent temporal averaging of $b_i(t)$ to obtain $f_i(t)$ constitutes a remapping of an amplitude-encoding quantity into a duty cycle-encoding quantity, mediated by threshold function $H(c_i - \Theta_i)$

- Different scales of $c_i(t)$ can result in similar T_1 segments depending on the magnitude of the derivative of $c_i(t)$ in temporal proximity to time points at which $c_i(t)$ crosses threshold value Θ_i

→ The steeper the slope of $c_i(t)$, the less T_1 changes with scale variations
→ If T_1 is invariant to scale variation in $c_i(t)$, then so is $f_i(t)$

- Suggests a relatively simple rule for optimal choice of threshold value Θ_i:
→ Find amplitude c_i that maximizes absolute derivative of $c_i(t)$ over time
→ Optimal with respect to intensity invariance of $f_i(t)$, not necessarily for other criteria such as song-noise separation or diversity between features

- Nonlinear operations can be used to detach representations from graded physical stimulus (to fasciliate categorical behavioral decision-making?):
1) Capture sufficiently precise amplitude information: $x_{\text{env}}(t)$, $x_{\text{adapt}}(t)$
→ Closely following the AM of the acoustic stimulus
2) Quantify relevant stimulus properties on a graded scale: $c_i(t)$
→ More decorrelated representation, compared to prior stages
3) Nonlinearity: Distinguish between "relevant vs irrelevant" values: $b_i(t)$
→ Trading a graded scale for two or more categorical states
4) Represent stimulus properties under relevance constraint: $f_i(t)$
→ Graded again but highly decorrelated from the acoustic stimulus
5) Categorical behavioral decision-making requires further nonlinearities
→ Parameters of a behavioral response may be graded (e.g. approach speed), initiation of one behavior over another is categorical (e.g. approach/stay)

4 Discriminating species-specific song patterns in feature space

5 Conclusions & outlook