Emergent intensity invariance vs. signal-to-noise ratio at three consecutive processing stages along the grasshopper song recognition pathway

Jona Hartling, Jan Benda

1 Exploring a grasshopper's sensory world

Our scientific understanding of sensory processing systems results from the distributed accumulation of anatomical, physiological and ethological evidence. This process is undoubtedly without alternative; however, it leaves us with the challenge of integrating the available fragments into a coherent whole in order to address issues such as the interaction between individual system components, the functional limitations of the system overall, or taxonomic comparisons between systems that process the same sensory modality. Any unified framework that captures the essential functional aspects of a given sensory system thus has the potential to deepen our current understanding and fasciliate systematic investigations. However, building such a framework is a challenging task. It requires a wealth of existing knowledge of the system and the signals it operates on, a clearly defined scope, and careful reduction, abstraction, and formalization of the underlying structures and mechanisms.

One sensory system about which extensive information has been gathered over the years is the auditory system of grasshoppers (*Acrididae*). Grasshoppers rely on their sense of hearing primarily for intraspecific communication, which includes mate attraction (D. v. Helversen 1972) and evaluation (Stange and Ronacher 2012), sender localization (D. v. Helversen and Rheinlaender 1988), courtship display (Elsner 1968), rival deterrence (Greenfield and Minckley 1993), and loss-of-signal predator alarm (SOURCE). In accordance with this rich behavioral repertoire, grasshoppers have evolved a variety of sound production mechanisms to generate acoustic communication signals for different contexts and ranges using their wings, hindlegs, or mandibles (Otte 1970). Among the most conspicuous acoustic signals of grasshoppers are their species-specific calling songs, which broadcast the presence of the singing individual — mostly the males of the species — to potential mates within range. These songs are usually more characteristic of a species

than morphological traits (Tishechkin and Vedenina 2016; Tarasova et al. 2021), which can vary greatly within species (Rowell 1972; Köhler et al. 2017). The reliance on songs to mediate reproduction represents a strong evolutionary driving force, that resulted in a massive species diversification (Vedenina and Mugue 2011; Sevastianov et al. 2023), with over 6800 recognized grasshopper species in the *Acrididae* family (Cigliano et al. 2024). It is this diversity of species, and the crucial role of acoustic communication in its emergence, that makes the grasshopper auditory system an intriguing candidate for attempting to construct a functional model framework. As a necessary reduction, the model we propose here focuses on the pathway responsible for the recognition of species-specific calling songs, disregarding other essential auditory functions such as directional hearing (D. v. Helversen 1984; Ronacher, D. v. Helversen, and Helversen 1986; D. v. Helversen and Rheinlaender 1988).

To understand the functional challenges faced by the grasshopper auditory system, one has to understand the properties of the songs it is designed to recognize. Grasshopper songs are amplitude-modulated broad-band acoustic signals. Most songs are produced by stridulation, during which the animal pulls the serrated stridulatory file on its hindlegs across a resonating vein on the forewings (O. v. Helversen and Elsner 1977; Stumpner and O. v. Helversen 1994; D. v. Helversen and O. v. Helversen 1997). Every tooth that strikes the vein generates a brief pulse of sound. Multiple pulses make up a syllable; and the alternation of syllables and relatively quiet pauses forms a characteristic, through noisy, waveform pattern. Song recognition depends on certain temporal and structural parameters of this pattern, such as the duration of syllables and pauses (D. v. Helversen 1972), the slope of pulse onsets (D. v. Helversen 1993), and the accentuation of syllable onsets relative to the preceeding pause (Balakrishnan et al. 2001; D. v. Helversen, Balakrishnan, and Helversen 2004). The amplitude modulation of the song is sufficient for recognition (D. v. Helversen and O. v. Helversen 1997). However, the essential recognition cues can vary considerably with external physical factors, which requires the auditory system to be invariant to such variations in order to reliably recognize songs under different conditions. For instance, the temporal structure of grasshopper songs warps with temperature (Skovmand and Boel Pedersen 1983). The auditory system can compensate for this variability by reading out relative temporal relationships rather than absolute time intervals (Creutzig, Wohlgemuth, et al. 2009; Creutzig, Benda, et al. 2010), as those remain relatively constant across different temperatures (D. v. Helversen 1972). Another, perhaps even more fundamental external source of song variability lays in the attenuation of sound intensity with increasing distance to the sender. Sound attenuation depends on both the frequency content of the signal and the vegetation of the habitat (Michelsen 1978). For the receiving auditory system, this has two major implications. First, the amplitude dynamics of the song pattern are steadily degraded over distance, which limits the effective com-

munication range of grasshoppers to 1 - 2 m in their typical grassland habitats (Lang 2000). Second, the overall intensity level of songs at the receiver's position varies depending on the location of the sender, which should ideally not affect the recognition of the song pattern. This neccessitates that the auditory system achieves a certain degree of intensity invariance — a time scale-selective sensitivity to faster amplitude dynamics and simultaneous insensitivity to slower, more sustained amplitude dynamics. Intensity invariance in different auditory systems is often associated with neuronal adaptation (Benda and Hennig 2008; Barbour 2011; Ozeri-Engelhard et al. 2018; more general: Benda 2021). In the grasshopper auditory system, a number of neuron types along the processing chain exhibit spike-frequency adaptation in response to sustained stimulus intensities (Römer 1976; Gollisch and Herz 2004; Hildebrandt et al. 2009; Clemens, Weschke, et al. 2010; Fisch et al. 2012) and thus likely contribute to the emergence of intensity-invariant song representations. This means that intensity invariance is not the result of a single processing step but rather a gradual process, in which different neuronal populations contribute to varying degrees (Clemens, Weschke, et al. 2010) and by different mechanisms (Hildebrandt et al. 2009). Approximating this process within a functional model framework thus requires a considerable amount of simplification. In this work, we demonstrate that even a small number of basic physiologically inspired signal transformations — specifically, pairs of nonlinear and linear operations — is sufficient to achieve a meaningful degree of intensity invariance.

Invariance to non-informative song variations is crucial for reliable song recognition; however, it is not sufficient to this end. In order to recognize a conspecific song as such, the auditory system needs to extract sufficiently informative features of the song pattern and then integrate the gathered information into a final categorical percept. Previous authors have proposed a functional model framework that describes this process — feature extraction, evidence accumulation, and categorical decision making — in both crickets (Clemens and Hennig 2013; Hennig et al. 2014) and grasshoppers (Clemens and Ronacher 2013; review on both: Ronacher, Hennig, and Clemens 2015). Their framework provides a comprehensible and biologically plausible account of the computational mechanisms required for species-specific song recognition, which has served as the inspiration for the development of the model pathway we propose here. The existing framework relies on pulse trains as input signals, which were designed to capture the essential structural properties of natural song envelopes (Clemens and Ronacher 2013). In the first step, a bank of parallel linear-nonlinear feature detectors is applied to the input signal. Each feature detector consists of a convolutional filter and a subsequent sigmoidal nonlinearity. The outputs of these feature detectors are temporally averaged to obtain a single feature value per detector, which is then assigned a specific weight. The linear combination of weighted feature values results in a single preference value, that serves as predictor for the behav-

ioral response of the animal to the presented input signal. Our model pathway adopts the general structure of the existing framework but modifies it in several key aspects. The convolutional filters, which have previously been fitted to behavioral data for each individual species (Clemens and Hennig 2013), are replaced by a larger, generic set of unfitted Gabor basis functions in order to cover a wide range of possible song features across different species. Gabor functions approximate the general structure of the filters used in the existing framework as well as the filter functions found in various auditory neurons (Rokem et al. 2006; Clemens, Kutzki, et al. 2011; Clemens, Wohlgemuth, and Ronacher 2012). The fitted sigmoidal nonlinearities in the existing framework consistently exhibited very steep slopes and are therefore replaced by shifted Heaviside step-functions, which results in a binarization of the feature detector outputs. Another, more substantial modification is that the feature detector outputs are temporally averaged in a way that does not condense them into single feature values but retains their time-varying structure. This is in line with the fact that songs are no discrete units but part of a continuous acoustic stream that the auditory system has to process in real time. Moreover, a time-varying feature representation only stabilizes after a certain delay following the onset of a song, which emphasizes the temporal dynamics of evidence accumulation towards a final categorical decision. The most notable difference between our model pathway and the existing framework, however, lays in the addition of a physiologically inspired preprocessing stage, whose starting point corresponds to the initial reception of airborne sound waves. This allows the model to operate on unmodified recordings of natural grasshopper songs instead of condensed pulse train approximations, which widens its scope towards more realistic, ecologically relevant scenarios. For instance, we were able to investigate the contribution of different processing stages to the emergence of intensity-invariant song representations based on actual field recordings of songs at different distances from the sender. In the following, we outline the structure of the proposed model of the grasshopper auditory pathway, from the initial reception of sound waves up to the generation of a high-dimensional, time-varying feature representation that is suitable for species-specific song recognition. We provide a side-by-side account of the known physiological processing steps and their functional approximation by basic mathematical operations. We then elaborate on two key mechanisms that drive the emergence of intensity-invariant song representations within the auditory pathway.

2 Methods

2.1 Functional model of the grasshopper song recognition pathway

The essence of constructing a functional model of a given system is to gain a sufficient understanding of the system's essential structural components and their presumed functional roles;

and to then build a formal framework of manageable complexity around these two aspects. Anatomically, the organization of the grasshopper song recognition pathway can be outlined as a feed-forward network of three consecutive neuronal populations (Fig. 1a-c): Peripheral auditory receptor neurons, whose axons enter the ventral nerve cord at the level of the metathoracic ganglion; local interneurons that remain exclusively within the thoracic region of the ventral nerve cord; and ascending neurons projecting from the thoracic region towards the supraesophageal ganglion (Rehbein et al. 1974; Rehbein 1976; Eichendorf and Kalmring 1980). The input to the network originates at the tympanal membrane, which acts as acoustic receiver and is coupled to the dendritic endings of the receptor neurons (Gray 1960). The outputs from the network converge in the supraesophageal ganglion, which is presumed to harbor the neuronal substrate for conspecific song recognition and response initiation (Ronacher, D. v. Helversen, and Helversen 1986; Bauer and Helversen 1987; Bhavsar et al. 2017). Functionally, the ascending neurons are the most diverse of the three populations along the pathway. Individual ascending neurons possess highly specific response properties that contrast with the rather homogeneous response properties of the preceding receptor neurons and local interneurons (Clemens, Kutzki, et al. 2011), indicating a transition from a uniform population-wide processing stream into several parallel branches. Based on these anatomical and physiological considerations, the overall structure of the model pathway is divided into two distinct stages (Fig. 1d). The preprocessing stage incorporates the known physiological processing steps at the levels of the tympanal membrane, the receptor neurons, and the local interneurons; and operates on one-dimensional signal representations. The feature extraction stage corresponds to the processing within the ascending neurons and further downstream towards the supraesophageal ganglion; and operates on high-dimensional signal representations. The details of each physiological processing step and its functional approximation within the two stages are outlined in the following sections.

2.1.1 Population-driven signal preprocessing

Grasshoppers receive airborne sound waves by a tympanal organ at either side of the body. The tympanal membrane acts as a mechanical resonance filter for sound-induced vibrations (Windmill et al. 200 Malkin et al. 2014). Vibrations that fall within specific frequency bands are focused on different membrane areas, while others are attenuated. This processing step can be approximated by an initial bandpass filter

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

applied to the acoustic input signal $x_{\text{raw}}(t)$. The auditory receptor neurons transduce the vibrations of the tympanal membrane into sequences of action potentials. Thereby, they encode the amplitude modulation, or envelope, of the signal (Machens, Prinz, et al. 2001), which likely

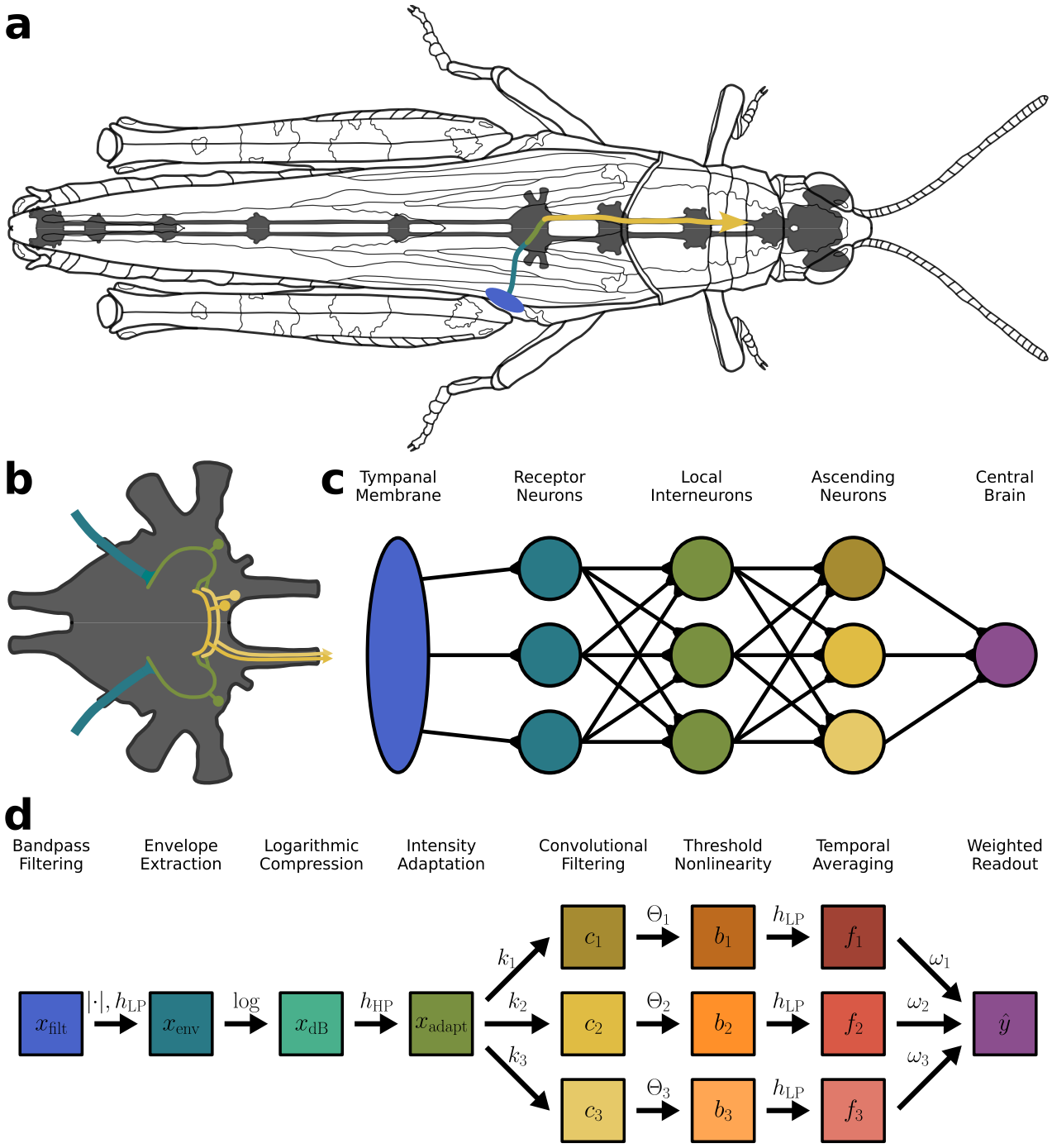

Fig. 1: Schematic organisation of the grasshopper song recognition pathway and structure of the functional model pathway. **a**: Simplified course of the pathway in the grasshopper, from the tympanal membrane over receptor neurons, local interneurons, and ascending neurons further towards the supraesophageal ganglion. **b**: Schematic of synaptic connections between the three neuronal populations within the metathoracic ganglion. **c**: Network representation of neuronal connectivity. **d**: Flow diagram of consecutive signal representations (boxes) and transformations (arrows) along the model pathway. All representations are time-varying. 1st half: Preprocessing stage (one-dimensional representation). 2nd half: Feature extraction stage (high-dimensional representation).

involves a rectifying nonlinearity (Machens, Stemmler, et al. 2001). This can be modelled as full-wave rectification followed by lowpass filtering

$$x_{\mathrm{env}}(t) \;=\; |x_{\mathrm{filt}}(t)| \;*\; h_{\mathrm{LP}}(t), \qquad f_{\mathrm{cut}} \;=\; 250\,\mathrm{Hz} \tag{2}$$

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

$$x_{\mathrm{log}}(t) \;=\; 20 \;\cdot\; \log_{10} \frac{x_{\mathrm{env}}(t)}{x_{\mathrm{ref}}}, \qquad x_{\mathrm{ref}} \;=\; 1 \tag{3}$$

relative to the common reference intensity x_{ref}. Both the receptor neurons (Römer 1976; Gollisch and Herz 2004; Fisch et al. 2012) and, on a larger scale, the subsequent local interneurons (Hildebrandt et al. 2009; Clemens, Weschke, et al. 2010) adapt their firing rates in response to sustained stimulus intensity levels, which allows for the robust encoding of faster amplitude modulations against a slowly changing overall baseline intensity. Functionally, the adaptation mechanism resembles a highpass filter

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

over the logarithmically scaled envelope $x_{\mathrm{log}}(t)$. This processing step concludes the preprocessing stage of the model pathway. The resulting intensity-adapted envelope $x_{\mathrm{adapt}}(t)$ is then passed on from the local interneurons to the ascending neurons, where it serves as the basis for the following feature extraction stage.

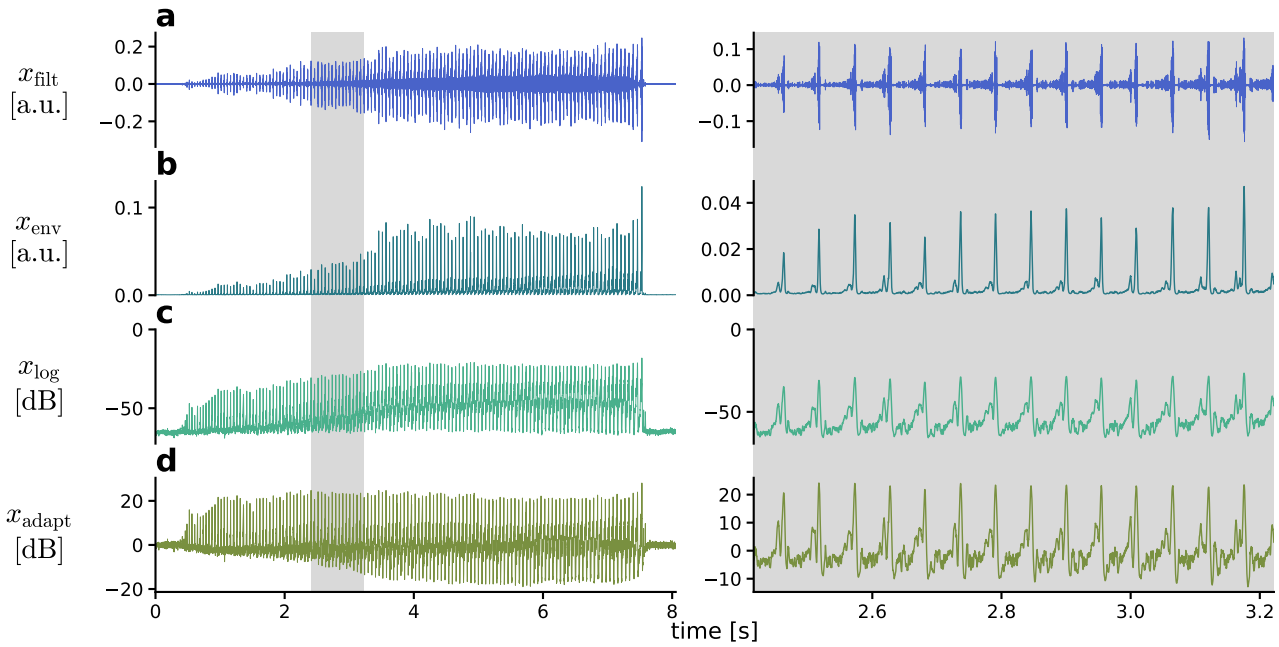

Fig. 2: Representations of a song of *O. rufipes* during the preprocessing stage. **a**: Bandpass filtered tympanal signal $x_{\text{filt}}(t)$. **b**: Signal envelope $x_{\text{env}}(t)$. **c**: Logarithmically compressed envelope $x_{\text{log}}(t)$. **d**: Intensity-adapted envelope $x_{\text{adapt}}(t)$.

2.1.2 Feature extraction by individual neurons

The ascending neurons extract and encode a number of different features of the preprocessed signal. As a population, they hence represent the signal in a higher-dimensional space than the preceding receptor neurons and local interneurons. Each ascending neuron is assumed to scan the signal for a specific template pattern, which can be thought of as a kernel of a particular structure and on a particular time scale. This process, known as template matching, can be modelled as a convolution

$$c_i(t) \;=\; x_{\text{adapt}}(t) \;*\; k_i(t) = \int_{-\infty}^{+\infty} x_{\text{adapt}}(\tau) \;\cdot\; k_i(t \;-\; \tau)\, d\tau \tag{5}$$

of the intensity-adapted envelope $x_{\text{adapt}}(t)$ with a kernel $k_i(t)$ per ascending neuron. We use Gabor kernels as basis functions for creating different template patterns. An arbitrary one-dimensional, real Gabor kernel is generated by multiplication of a Gaussian envelope and a sinusoidal carrier

$$k_i(t,\, \sigma_i,\, \omega_i,\, \phi_i) \;=\; e^{-\frac{t^2}{2{\sigma_i}^2}} \;\cdot\; \sin(\omega_i\, t \;+\; \phi_i), \qquad \omega_i \;=\; 2\pi f_{\text{sin}} \tag{6}$$

with Gaussian standard deviation or kernel width σ_i, carrier frequency ω_i, and carrier phase ϕ_i. Different combinations of σ and ω result in Gabor kernels with different lobe number n,

which is the number of half-periods of the carrier that fit under the Gaussian envelope within reasonable limits of attenuation. The interval under the Gaussian envelope that contains the relevant lobes of the kernel can be defined as Gaussian full-width measured at relative peak height h_{rel}

$$\mathrm{FWRH}(\sigma,\, h_{\mathrm{rel}}) \,=\, 2 \,\cdot\, \sqrt{-2 \,\cdot\, \ln h_{\mathrm{rel}}} \cdot\, \sigma, \qquad h_{\mathrm{rel}} \,\in\, (0,\, 1] \tag{7}$$

With this, an appropriate carrier frequency ω for obtaining a Gabor kernel with width σ and desired lobe number n can be approximated as

$$\omega(n,\, \sigma,\, h_{\mathrm{rel}}) \,=\, \frac{n \,+\, \beta_0}{4 \,\cdot\, \sqrt{-2 \,\cdot\, \ln h_{\mathrm{rel}}}}, \qquad n \,\geq\, 2 \;\forall\; n \,\in\, \mathbb{Z} \tag{8}$$

where β_0 is a small positive offset to the near-linear relationship between ω and n to balance the amplitude of the n desired lobes of the kernel — which should be maximized — against the amplitude of the next-outer lobes, which should not exceed the threshold value determined by h_{rel}. For $n = 1$, carrier frequency ω is set to zero, which results in a simple Gaussian kernel. Carrier phase ϕ determines the position of the kernel lobes relative to the kernel center. By setting ϕ to one of only four specific phase values (Tab. 1), we restrict the Gabor kernels to be either even functions (mirror-symmetric, uneven n) or odd functions (point-symmetric, even n) with either positive or negative sign, which refers to the sign of the kernel's central lobe (even kernels) or the left of the two central lobes (odd kernels).

Tab. 1: Values of phase ϕ that are specific for the four major groups of Gabor kernels.

sign	even kernels	odd kernels
$+$	$+\pi\,/\,2$	π
$-$	$-\pi\,/\,2$	0

These four major groups of Gabor kernels allow for the extraction of different types of signal features, such as the presence of peaks (even, +), troughs (even, −), onsets (odd, +), and offsets (odd, −) at various time scales. Following the convolutional template matching, each kernel-specific response $c_i(t)$ is passed through a shifted Heaviside step-function $H(c_i \,-\, \Theta_i)$ with threshold value Θ_i to obtain a binary response

$$b_i(t,\, \Theta_i) \,=\, \begin{cases} 1, & c_i(t) \,>\, \Theta_i \\ 0, & c_i(t) \,\leq\, \Theta_i \end{cases} \tag{9}$$

which can be thought of as a categorization into "relevant" and "irrelevant" response values. In the grasshopper, these thresholding nonlinearities might either be part of the processing within the ascending neurons or take place further downstream (SOURCE). Finally, the responses of

the ascending neurons are assumed to be integrated somewhere in the supraesophageal ganglion (Ronacher, D. v. Helversen, and Helversen 1986; Bauer and Helversen 1987; Bhavsar et al. 2017). This processing step can be approximated as temporal averaging of the binary responses $b_i(t)$ by a lowpass filter

$$f_i(t) \;=\; b_i(t) \;*\; h_{\mathrm{LP}}(t), \qquad f_{\mathrm{cut}} \;=\; 1\,\mathrm{Hz} \tag{10}$$

to obtain a final set of slowly changing kernel-specific features $f_i(t)$. In the resulting high-dimensional feature space, different species-specific song patterns are characterized by a distinct combination of feature values, which can be read out by a simple linear classifier.

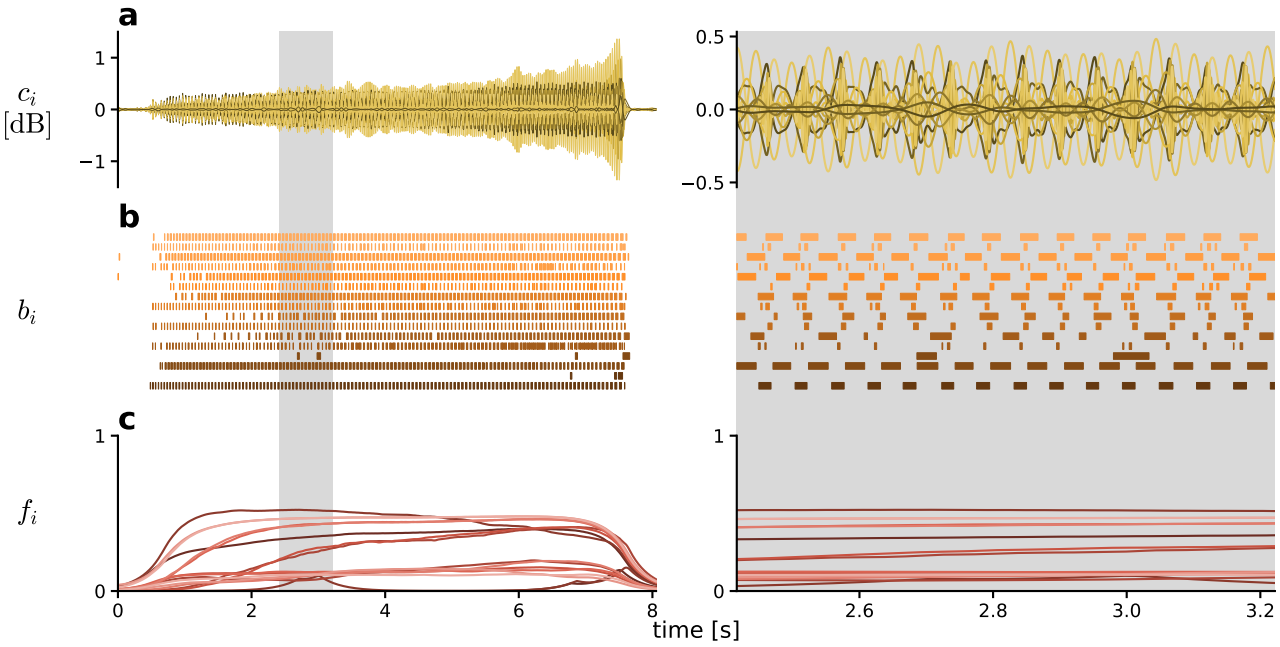

Fig. 3: Representations of a song of *O. rufipes* during the feature extraction stage. Different color shades indicate different types of Gabor kernels with specific lobe number n and either + or − sign, sorted (dark to light) first by increasing n and then by sign ($1 \leq n \leq 4$; first +, then − for each n; two kernel widths σ of 4 ms and 32 ms per type; 8 types, 16 kernels in total). **a**: Kernel-specific filter responses $c_i(t)$. **b**: Binary responses $b_i(t)$. **c**: Finalized features $f_i(t)$.

2.2 Simulation-based analysis of the model pathway

2.2.1 Data sourcing

All simulations were based on a dataset that was assembled from five different sources, each of which is an established reference for the identification of European grasshopper species. The dataset was limited to six species from the species-rich *Gomphocerinae* sub-family that are known to be common throughout Central and Southern Europe. All recordings were converted to standard *.wav* format with a sampling rate of 44.1 kHz and an amplitude scale in arbitrary units. Individual songs were then cut from each recording. The dataset includes a total of 31

recordings across species, which amounts to a total of 153 isolated songs. However, the number of available species-specific songs varies greatly across species, with a maximum of 48 songs for *C. biguttulus* and a minimum of 6 songs for *C. mollis* (Tab. 2).

- "Heuschrecken beobachten, bestimmen" by Heiko Bellmann
  1st edition, 1993, Naturbuch, Augsburg
- "Gesänge der heimischen Heuschrecken. Akustisch-optische Bestimmungshilfe."
  by Karl-Heinz Garberding, Deutscher Jugendbund für Naturbeobachtung
  1st edition, 2001, DJN, Göttingen
- "Heuschrecken – Die Stimmen von 61 heimischen Arten" by Heiko Bellmann
  1st edition, 2004, AMPLE, Germering
- "Fauna d'Italia XLVIII – Orthoptera" by Bruno Massa, Paolo Fontana, Filippo M. Buzzetti, Roy M.J.C. Kleukers, Baudewijn Odé
  1st edition, 2012, edagricola, Milano
- "Singing Orthoptera of Slovenia" by Stanislav Gomboc, Blaz Segula
  1st edition, 2014, EGEA, Ljubljana

Tab. 2: Overview of the six grasshopper species from the *Gomphocerinae* sub-family, the number of sources per species, the number of available recordings across sources, and the number of isolated songs across recordings.

Species	**Sources**	**Recordings**	**Songs**
Chorthippus biguttulus	5	6	48
Chorthippus mollis	3	3	6
Chrysochraon dispar	4	5	45
Gomphocerippus rufus	4	8	16
Omocestus rufipes	4	5	14
Pseudochorthippus parallelus	4	4	24

2.2.2 Generating synthetic input signals

Different processing steps along the model pathway were tested for intensity invariance by generating synthetic input signals $x(t)$ of varying intensity, transforming them through the respective processing steps, and comparing the resulting signal representations. Inputs were generated for two distinct cases. In the idealized, noiseless case, $x(t)$ consists of a song component $s(t)$ with $\sigma_s = 1$ and a multiplicative scale α:

$$x(t) = \alpha \cdot s(t), \qquad \alpha \geq 0 \tag{11}$$

In the noiseless case, $x(t)$ is hence only a scaled version of $s(t)$ with $\sigma_x = \alpha$. In the more realistic, noisy case, $x(t)$ consists of the same song component $s(t)$ scaled by α and an additive noise component $\eta(t)$ with $\sigma_\eta = 1$:

$$x(t) = \alpha \cdot s(t) + \eta(t), \qquad \alpha \geq 0 \tag{12}$$

Accordingly, the SNR of input $x(t)$ in the noisy case equals the squared α value:

$$\mathrm{SNR}_x(\alpha) = \frac{(\alpha \cdot \sigma_\mathrm{s})^2}{\sigma_\eta^2} = \alpha^2, \qquad \sigma_\mathrm{s} = \sigma_\eta = 1 \tag{13}$$

For most analyses, it would be sufficient if input $x(t)$ corresponds to the signal representation immediately before the first of the tested transformations. For instance, when testing the effects of logarithmic compression (Eq. 3), $x(t)$ would correspond to the signal envelope $x_\mathrm{env}(t)$. However, in this particular case, $x_\mathrm{env}(t)$ results from a nonlinear transformation (Eq. 2), which cannot be synthesized as an additive mixture of $s(t)$ and $\eta(t)$. For this reason, any input $x(t)$ across all analyses corresponds not to the representation immediately before the tested transformations but its predecessor representation instead. Therefore, when testing logarithmic compression, $x(t)$ corresponds to the tympanal signal $x_\mathrm{filt}(t)$ instead of $x_\mathrm{env}(t)$.

The raw $s(t)$ was drawn from the dataset of isolated species-specific song recordings, whereas the raw $\eta(t)$ consists of a segment of normally distributed white noise. Both $s(t)$ and $\eta(t)$ were normalized to unit standard deviation. These can be used without further processing for all analyses where input $x(t)$ corresponds to $x_\mathrm{raw}(t)$. For analyses where $x(t)$ corresponds to a later representation, $s(t)$ and $\eta(t)$ were first processed along the model pathway up to the required representation, again normalized to unit standard deviation, and then used to generate $x(t)$ according to either Eq. 11 in the noiseless case or Eq. 12 in the noisy case.

2.2.3 Quantifying signal intensity across representations

All intensity measures were calculated over a manually labeled segment within each song. Segments always excluded the first and last few syllables to allow slowly changing representations such as $f_i(t)$ to stabilize. The duration of each segment and the number of contained syllables depends on the duration of the species-specific song. Care was taken to ensure that the segment contained a sufficient number of syllables to obtain a reliable estimate of the intensity measures.

The standard deviation σ was used as a measure of intensity for all representations resulting from the transformation of input $x(t)$ up to and including the kernel responses $c_i(t)$, for which individual σ_{c_i} were used as kernel-specific intensity measures. The binary responses $b_i(t)$ were deemed to similar to the features $f_i(t)$ to warrant their own intensity measure and were hence

omitted from all related analyses. For $f_i(t)$, σ is not an appropriate intensity measure because each $f_i(t)$ is ideally constant with $\sigma = 0$ for the duration of a song. Therefore, the average value μ_{f_i} of each $f_i(t)$ was used as a kernel-specific intensity measure instead.

It is arguably not ideal to quantify the intensity of $c_i(t)$ and $f_i(t)$ separately for each kernel. Overall, these representations are not separate signals bundled together but rather a set that acts as a unit with a single intensity measure. However, there is no straightforward way to quantify the intensity of $c_i(t)$ or $f_i(t)$ as a whole that would not entail a certain ambiguity, e. g by averaging across kernels. In this sense, we opted for the kernel-specific approach because it allows to asses differences in the dependency on α between individual members of either $c_i(t)$ and $f_i(t)$.

The absolute intensity measures allow to compare the intensity of a representation across different α values. Additionally, ratios were calculated between the intensity measures for $\alpha > 0$ and the respective pure-noise reference measure for $\alpha = 0$ to better compare the intensities of different representations. This is only possible in the noisy case, where input $x(t) = \eta(t)$ for $\alpha = 0$, whereas $x(t) = 0$ for $\alpha = 0$ in the noiseless case. At the level of input $x(t)$, the ratio of intensity measures depends on the square root of α:

$$\frac{\sigma_x}{\sigma_\eta} = \sqrt{\frac{\sigma_x^2}{\sigma_\eta^2}} = \sqrt{\frac{(\alpha \cdot \sigma_\mathrm{s})^2 + \sigma_\eta^2}{\sigma_\eta^2}} = \sqrt{\alpha^2 + 1}, \qquad \sigma_\mathrm{s} = \sigma_\eta = 1 \tag{14}$$

This holds only if $s(t) \perp \eta(t)$, so that $\sigma_x^2 = \sigma_\mathrm{s}^2 + \sigma_\eta^2$, which is a reasonable assumption for the raw $s(t)$ and $\eta(t)$. However, the dependency of the ratio on α is not necessarily the same for representations that are transformed from $x(t)$ by nonlinear operations, since these change the relationship of $s(t)$ and $\eta(t)$ in an unpredictable fashion. Furthermore, the ratio is not a proper SNR of the representation because it does not relate $s(t)$ to $\eta(t)$ within the representation but rather the entire representation to $\eta(t)$ alone. However, it still provides a useful measure of the relative intensity of a representation with and without $s(t)$, which is the closest we can get to the SNR of the representation. As such, the ratio of intensity measures is referred to as SNR in the following.

2.3 Field data-based analysis of the model pathway

3 Results

3.1 Mechanisms driving the emergence of intensity invariance

The robustness of song recognition is tied to the degree of intensity invariance of the finalized feature representation. Ideally, the values of each feature should depend only on the relative amplitude dynamics of the song pattern but not on the overall intensity of the song. In the grasshopper, the emergence of intensity-invariant representations along the song recognition pathway likely is a distributed process that involves different neuronal populations, which raises the question of what the essential computational mechanisms are that drive this process. Within the model pathway, we identified two key mechanisms that render the song representation more invariant to intensity variations. The two mechanisms each comprise a nonlinear signal transformation followed by a linear signal transformation but differ in the specific operations involved, as outlined in the following sections.

3.1.1 Full-wave rectification & lowpass filtering

The first nonlinear transformation along the model pathway is the full-wave rectification of the tympanal signal $x_{\mathrm{filt}}(t)$ during the extraction of the signal envelope (Eq. 2). Rectification transforms the distribution of $x_{\mathrm{filt}}(t)$ from an approximately zero-centered distribution with both positive and negative values into a strictly non-negative distribution. Signal envelope $x_{\mathrm{env}}(t)$ is then obtained by lowpass filtering the rectified $x_{\mathrm{filt}}(t)$. The effects of this transformation pair on SNR and potential intensity invariance were analyzed by rescaling and processing the input signal $x_{\mathrm{raw}}(t)$ and comparing standard deviations between the resulting $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$, once for the noiseless case (Fig. 4a) and once for the noisy case (Fig. 4b). In addition, the cutoff frequency f_{cut} of the lowpass filter was varied to investigate the influence of different filter bandwidths. In the noiseless case, the standard deviations of $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$ are each reduced compared to the input $x_{\mathrm{raw}}(t)$ by a multiplicative factor. These factors are constant across all α, which results in a downward shift of the respective curve on a double-logarithmic scale, away from the diagonal (Fig. 4c). For $x_{\mathrm{filt}}(t)$, the reduction is a consequence of the bandpass filtering (Eq. 1) of $x_{\mathrm{raw}}(t)$. For $x_{\mathrm{env}}(t)$, the standard deviation is further reduced compared to $x_{\mathrm{filt}}(t)$. Rectification contributes much less to this reduction than lowpass filtering. The degree of reduction by lowpass filtering depends on the cutoff frequency f_{cut}, with lower f_{cut} (narrow bandwidth) resulting in a stronger reduction. In the noisy case, the standard deviations of $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$ can be related to the respective pure-noise reference standard deviation (Fig. 4d). This causes each curve to start with a constant regime of SNR

values near 1 for smaller α, which reflects the dominance of the noise component $\eta(t)$ over the song component $s(t)$ in the input $x_{\mathrm{raw}}(t)$. For larger α, all curves transition into a regime of linearly increasing SNR on a double-logarithmic scale. For $x_{\mathrm{filt}}(t)$, the linear part of the curve deviates only slightly from the diagonal. For $x_{\mathrm{env}}(t)$, however, the transition occurs at lower α compared to $x_{\mathrm{filt}}(t)$, and the linear part of the curve is shifted leftward away from the diagonal, which means that higher SNR values are achieved for the same α. This effect is more pronounced for lower f_{cut} of the lowpass filter and is presumably caused by the attenuation of high-frequency components in the signal, which are more prominent in the noise component $\eta(t)$ than in the song component $s(t)$. The effect also appears relatively consistent across different species, although small variations exist (Fig. 4e) that are presumably based on different song structures and frequency spectra. In summary, the standard deviation of $x_{\mathrm{env}}(t)$ has never been observed to transition into a saturation regime for larger α but rather continues to increase proportionally to α for all tested f_{cut}, in both the noiseless and the noisy case and across different species. Consequently, the combination of rectification and lowpass filtering does not contribute to intensity invariance. However, this transformation pair does improve the SNR of $x_{\mathrm{env}}(t)$ relative to $x_{\mathrm{filt}}(t)$ and thus provides subsequent processing stages with a more robust input representation and higher input SNR.

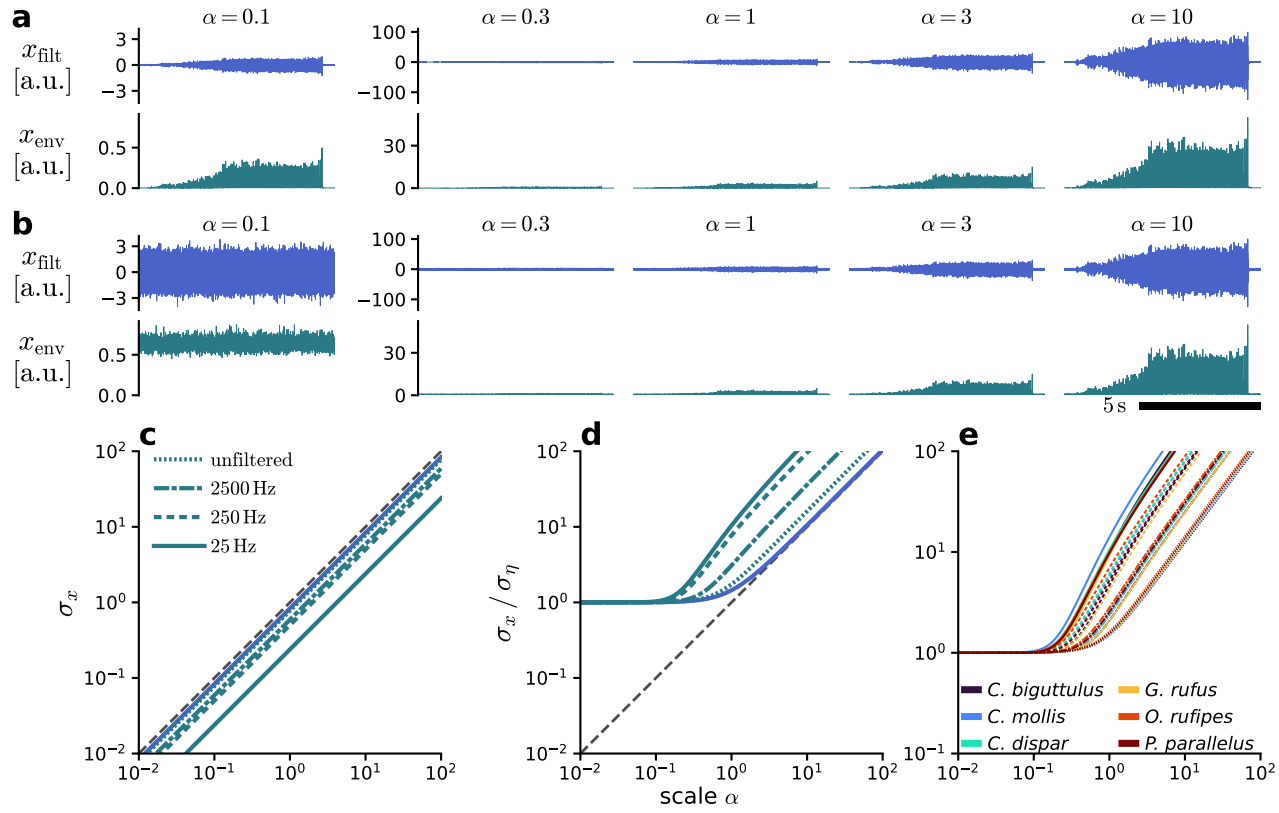

Fig. 4: Rectification and lowpass filtering improves SNR but does not contribute to intensity invariance. Input $x_{\mathrm{raw}}(t)$ consists of song component $s(t)$ scaled by α with optional noise component $\eta(t)$ and is successively transformed into tympanal signal $x_{\mathrm{filt}}(t)$ and envelope $x_{\mathrm{env}}(t)$. Different line styles indicate different cutoff frequencies f_{cut} of the lowpass filter extracting $x_{\mathrm{env}}(t)$. **Top**: Example representations of $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$ for different α. **a**: Noiseless case. **b**: Noisy case. **Bottom**: Intensity metrics over a range of α. **c**: Noiseless case: Standard deviations σ_x of $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$. **d**: Noisy case: Ratios of σ_x of $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$ to the respective reference standard deviation σ_η for input $x_{\mathrm{raw}}(t) = \eta(t)$. **e**: Ratios of σ_x to σ_η of $x_{\mathrm{env}}(t)$ as in **d** for different species (averaged over songs and recordings, see appendix Fig. 13).

3.1.2 Logarithmic compression & spike-frequency adaptation

The second nonlinear transformation along the model pathway is the logarithmic compression of the signal envelope $x_{\mathrm{env}}(t)$ into $x_{\log}(t)$, Eq. 3, which is then followed by the highpass filtering of $x_{\log}(t)$, Eq. 4, to obtain the intensity-adapted envelope $x_{\mathrm{adapt}}(t)$. The interplay of this transformation pair was analyzed by rescaling and processing the input signal $x_{\mathrm{filt}}(t)$ and comparing standard deviations between the resulting $x_{\mathrm{env}}(t)$, $x_{\log}(t)$, and $x_{\mathrm{adapt}}(t)$. It is necessary to use $x_{\mathrm{filt}}(t)$ as input for this analysis instead of $x_{\mathrm{env}}(t)$, because $x_{\mathrm{env}}(t)$ results from a nonlinear transformation and hence cannot be synthesized as an additive mixture of song component $s(t)$ and noise component $\eta(t)$. However, it is much easier to conceive a mathematical description of the effects of logarithmic compression and adaptation if $x_{\mathrm{env}}(t)$ itself is assumed to be composed

of $s(t)$ and $\eta(t)$. In the noiseless case (Fig. 5a), $x_{\mathrm{env}}(t)$ takes the form of

$$x_{\mathrm{env}}(t) \;=\; \alpha \,\cdot\, s(t), \qquad x_{\mathrm{env}}(t) \;>\; 0 \;\forall\; t \;\in\; \mathbb{R} \tag{15}$$

The standard deviation of $x_{\mathrm{env}}(t)$ increases linearly with α on a double-logarithmic scale and is slightly reduced (Fig. 5c) compared to the input $x_{\mathrm{filt}}(t)$, which is consistent with the results of the previous analysis (Fig. 4c). By conversion of $x_{\mathrm{env}}(t)$ to decibel scale, α turns from a multiplicative scale in linear space into an additive term, or offset, in logarithmic space:

$$x_{\log}(t) \;=\; 20 \,\cdot\, \log_{10}\left[\,\alpha \,\cdot\, s(t)\,\right] \;=\; 20 \,\cdot\, \left[\log_{10}\alpha \;+\; \log_{10} s(t)\right], \qquad \alpha \;>\; 0 \tag{16}$$

The highpass filtering of $x_{\log}(t)$ can be approximated as a subtraction of the local signal offset within a suitable time interval $0 \ll T_{\mathrm{HP}} < \frac{1}{f_{\mathrm{cut}}}$:

$$x_{\mathrm{adapt}}(t) \;\approx\; x_{\log}(t) \;-\; 20 \,\cdot\, \log_{10}\alpha \;=\; 20 \,\cdot\, \log_{10} s(t) \tag{17}$$

This eliminates α from $x_{\mathrm{adapt}}(t)$ and thus renders it perfectly intensity-invariant, with a constant standard deviation of around 10 dB across all $\alpha > 0$ (Fig. 5c). In contrast, in the noisy case (Fig. 5b), $x_{\mathrm{env}}(t)$ takes the form of

$$x_{\mathrm{env}}(t) \;=\; \alpha \,\cdot\, s(t) \;+\; \eta(t), \qquad x_{\mathrm{env}}(t) \;>\; 0 \;\forall\; t \;\in\; \mathbb{R} \tag{18}$$

Similar to the previous analysis (Fig. 4d), the ratio of the standard deviation of $x_{\mathrm{env}}(t)$ to its pure-noise reference standard deviation on a double-logarithmic scale follows a constant regime for small α and a linearly increasing regime for larger α (Fig. 5d). Decibel conversion of $x_{\mathrm{env}}(t)$

$$x_{\log}(t) \;=\; 20 \,\cdot\, \left(\log_{10}\alpha \;+\; \log_{10}\left[s(t) \;+\; \frac{\eta(t)}{\alpha}\right]\right), \qquad \alpha \;>\; 0 \tag{19}$$

allows for the separation of α from $s(t)$ but introduces a scaling of $\eta(t)$ by the inverse of α, which remains present even after the offset subtraction:

$$x_{\mathrm{adapt}}(t) \;\approx\; 20 \,\cdot\, \log_{10}\left[s(t) \;+\; \frac{\eta(t)}{\alpha}\right] \tag{20}$$

This means that, in the noisy case, α cannot be entirely eliminated from $x_{\mathrm{adapt}}(t)$, only redistributed between $s(t)$ and $\eta(t)$. If α is sufficiently large ($\alpha \gg 1$, saturation regime), $\eta(t)$ is attenuated to the point of being negligible, so that $x_{\mathrm{adapt}}(t)$ is a scale-free representation of $s(t)$. If α and $\eta(t)$ are at similar scales ($\alpha \approx 1$, transient regime), $x_{\mathrm{adapt}}(t)$ largely resembles $x_{\log}(t)$. Finally, if α is sufficiently small ($0 < \alpha \ll 1$, noise regime), $\eta(t)$ masks $s(t)$ even after the inten-

sity adaptation. Accordingly, the effective intensity invariance of $x_{\mathrm{adapt}}(t)$ through logarithmic compression and adaptation is limited by the SNR of $x_{\mathrm{env}}(t)$: Songs that have already sunken into the noise floor at the level of $x_{\mathrm{env}}(t)$ cannot be recovered by subsequent processing steps, which emphasizes the importance of the SNR improvement by rectification and lowpass filtering during the previous processing step (Fig. 4d). The general pattern of noise regime, transient regime, and saturation regime remains consistent across different species (Fig. 5e). However, the specific value of α at which the saturation regime is reached (see appendix Fig. 15) and the maximum SNR value of $x_{\mathrm{adapt}}(t)$ within the saturation regime vary considerably between and within species. For example, *C. biguttulus* and *C. mollis* display a noticably lower maximum SNR of $x_{\mathrm{adapt}}(t)$ compared to other species. These differences are not to be underestimated, since the SNR of $x_{\mathrm{adapt}}(t)$ within the saturation regime determines the maximum input SNR for subsequent processing steps. In other words, the fact that $x_{\mathrm{adapt}}(t)$ eventually reaches a saturation regime is, of course, desirable in the context of intensity invariance, but it also means to pass up on the higher SNR values that are achieved by $x_{\mathrm{env}}(t)$ for the same α (up to several orders of magnitude, Fig. 5d). This trade-off between intensity invariance and SNR is a recurring phenomenon that is further addressed in the following sections.

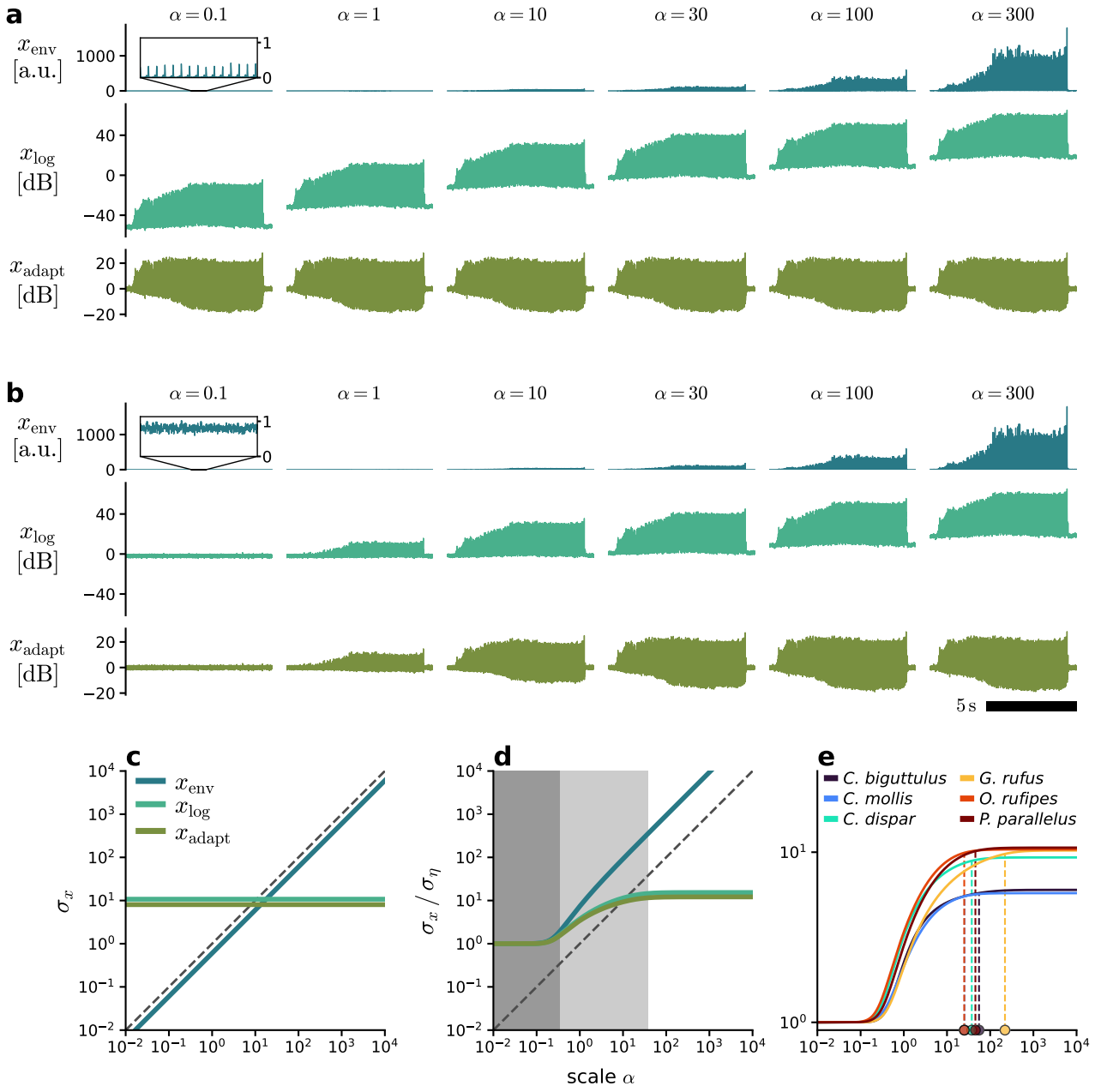

Fig. 5: Intensity invariance through logarithmic compression and adaptation is restricted by the noise floor and decreases SNR. Input $x_{\mathrm{filt}}(t)$ consists of song component $s(t)$ scaled by α with optional noise component $\eta(t)$ and is successively transformed into envelope $x_{\mathrm{env}}(t)$, logarithmically compressed envelope $x_{\mathrm{log}}(t)$, and intensity-adapted envelope $x_{\mathrm{adapt}}(t)$. **Top**: Example representations of $x_{\mathrm{env}}(t)$, $x_{\mathrm{log}}(t)$, and $x_{\mathrm{adapt}}(t)$ for different α. **a**: Noiseless case. **b**: Noisy case. **Bottom**: Intensity metrics over a range of α. **c**: Noiseless case: Standard deviations σ_x of $x_{\mathrm{env}}(t)$, $x_{\mathrm{log}}(t)$, and $x_{\mathrm{adapt}}(t)$. **d**: Noisy case: Ratios of σ_x of $x_{\mathrm{env}}(t)$, $x_{\mathrm{log}}(t)$, and $x_{\mathrm{adapt}}(t)$ to the respective reference standard deviation σ_η for input $x_{\mathrm{filt}}(t) = \eta(t)$. Shaded areas indicate 5 % (dark grey) and 95 % (light grey) curve span for $x_{\mathrm{adapt}}(t)$. **e**: Ratios of σ_x to σ_η of $x_{\mathrm{adapt}}(t)$ as in **d** for different species (averaged over songs and recordings, see appendix Fig 14). Dots indicate 95 % curve span per species.

3.1.3 Thresholding nonlinearity & temporal averaging

The third nonlinear transformation along the model pathway is the thresholding nonlinearity $H(c_i - \Theta_i)$ that transforms each kernel response $c_i(t)$ into a binary binary response $b_i(t)$, Eq. 9. This transformation takes place after the convolutional filtering of $x_{\mathrm{adapt}}(t)$ with kernel $k_i(t)$, Eq. 5, and is followed by the temporal averaging of $b_i(t)$ into the feature set $f_i(t)$ by a lowpass filter, Eq. 10. The effects of thresholding and temporal averaging are best illustrated based on a single kernel (Fig. 6) instead of the full set. For this analysis, input $x_{\mathrm{adapt}}(t)$ was rescaled (Fig. 6a) and convolved with kernel $k(t)$. The resulting kernel response $c(t)$ was passed through $H(c - \Theta)$ with three different threshold values Θ (Fig. 6b-d). Each resulting binary response $b(t)$ was transformed into $f(t)$, whose average feature value μ_f serves as a measure of intensity (Fig. 6ef). The thresholding nonlinearity $H(c - \Theta)$ categorizes the values of $c(t)$ into "relevant" ($c(t) > \Theta$, $b(t) = 1$) and "irrelevant" ($c(t) \leq \Theta$, $b(t) = 0$) response values. It thereby splits the probability density $p(c, T)$ of $c(t)$ within some observed time interval T into two complementary parts around Θ:

$$\int_{\Theta}^{+\infty} p(c, T)\, dc \;=\; 1 \;-\; \int_{-\infty}^{\Theta} p(c, T)\, dc \;=\; \frac{T_1}{T}, \qquad \int_{-\infty}^{+\infty} p(c, T)\, dc \;=\; 1 \tag{21}$$

The right-sided part of the split $p(c, T)$ corresponds to time T_1 where $c(t) > \Theta$, while the left-sided part corresponds to time $T_0 = T - T_1$ where $c(t) \leq \Theta$. The semi-definite integral over the right-sided part of $p(c, T)$ represents the ratio of time T_1 to total time T because the indefinite integral of a probability density is normalized to 1. The lowpass filtering of $b(t)$ can be approximated as temporal averaging over a suitable time interval $T_{\mathrm{LP}} > \frac{1}{f_{\mathrm{cut}}}$ in order to express $f(t)$ as a similar temporal ratio

$$f(t) \;\approx\; \frac{1}{T_{\mathrm{LP}}} \int_{t}^{t+T_{\mathrm{LP}}} b(\tau)\, d\tau \;=\; \frac{T_1}{T_{\mathrm{LP}}}, \qquad b(t) \;\in\; \{0,\, 1\} \tag{22}$$

of time T_1 during which $b(t)$ is 1 within the averaging interval T_{LP}. Therefore, the value of $f(t)$ at every time point t approximately signifies the cumulative probability that $c(t)$ exceeds Θ during the corresponding averaging interval T_{LP}:

$$f(t) \;\approx\; \int_{\Theta}^{+\infty} p(c,\, T_{\mathrm{LP}})\, dc \;=\; P(c \,>\, \Theta,\, T_{\mathrm{LP}}) \tag{23}$$

In a sense, $f(t)$ can be interpreted as some sort of duty cycle with respect to Θ. For example, a feature value of $f(t) = 0.4$ means that $c(t)$ exceeds Θ for approximately 40 % of the time within T_{LP} around t. In the most extreme cases, Θ lays either above the maximum of $c(t)$ or below the minimum of $c(t)$, which results in a minimum or maximum possible feature value of

$f(t) = 0$ (Fig. 6d, left column) or $f(t) = 1$, respectively.

Importantly, $f(t)$ neither retains information about the timing of individual threshold crossings nor the precise values of $c(t)$ apart from their relation to Θ. Accordingly, for a given Θ, different α can still result in similar T_1 segments (and hence similar feature values) depending on the magnitude of the derivative of $c(t)$ in temporal proximity to time points at which $c(t)$ crosses Θ: The steeper the slope of $c(t)$, the less T_1 changes with variations in α. The most reliable way of exploiting this invariant porperty of $f(t)$ is to set Θ to a value near 0, because these values are least affected by different scales of $c(t)$. For sufficiently large α, $f(t)$ then approaches the same constant μ_f in both the noiseless and the noisy case (Fig. 6e, saturation regime).

The value of μ_f in the saturation regime is independent of the precise value of Θ, but the value of α at which the saturation regime is reached decreses with Θ (Fig. 6e). Therefore, a threshold value of $\Theta = 0$ would be the optimal choice for achieving intensity invariance at the lowest possible α. In stark contrast, the closer Θ is to 0, the higher μ_f in response to the pure noise component $\eta(t)$ and the lower the resulting SNR of $f(t)$ between noise regime and saturation regime (Fig. 6b-d, left column, and Fig. 6e). It is even possible to achieve an "unlimited" SNR of $f(t)$ by setting Θ above the maximum of the pure-noise $c(t)$, so that any $\mu_f > 0$ indicates the presence of the song component $s(t)$ in input $x_{\text{adapt}}(t)$ at the cost of requiring a higher α to reach the saturation regime. This trade-off between intensity invariance and SNR has already been observed during the previous analysis on logarithmic compression and adaptation (Fig. 5d). However, the parameters that determine the SNR of $x_{\text{adapt}}(t)$ are much less understood and likely relate to properties of the signal, whereas the SNR of $f(t)$ depends on the choice of Θ and can be more directly manipulated by the system.

Finally, the effects of thresholding and temporal averaging must be seen in the context of the previous transformation pair of logarithmic compression and adaptation: In the current analysis, the input $x_{\text{adapt}}(t)$ can be rescaled by arbitrarily large α, while in the full pathway, the current input $x_{\text{adapt}}(t)$ is the output $x_{\text{adapt}}(t)$ of the previous transformation pair and is hence capped to a maximum standard deviation of around 10 dB (Fig. 5cd). This can be illustrated by plotting μ_f not over α (Fig. 6e) but over the standard deviation of input $x_{\text{adapt}}(t)$ instead (Fig. 6f). It becomes apparent that μ_f saturates only for standard deviations of $x_{\text{adapt}}(t)$ that would already be capped. Accordingly, $f(t)$ never reaches the saturation regime as determined by the current transformation pair but rather adheres to the saturation regime determined by the previous transformation pair. In this case, the saturated μ_f is not independent of Θ anymore. The consequences of this interaction between the two mechanisms of intensity invariance are further explored in a later section.

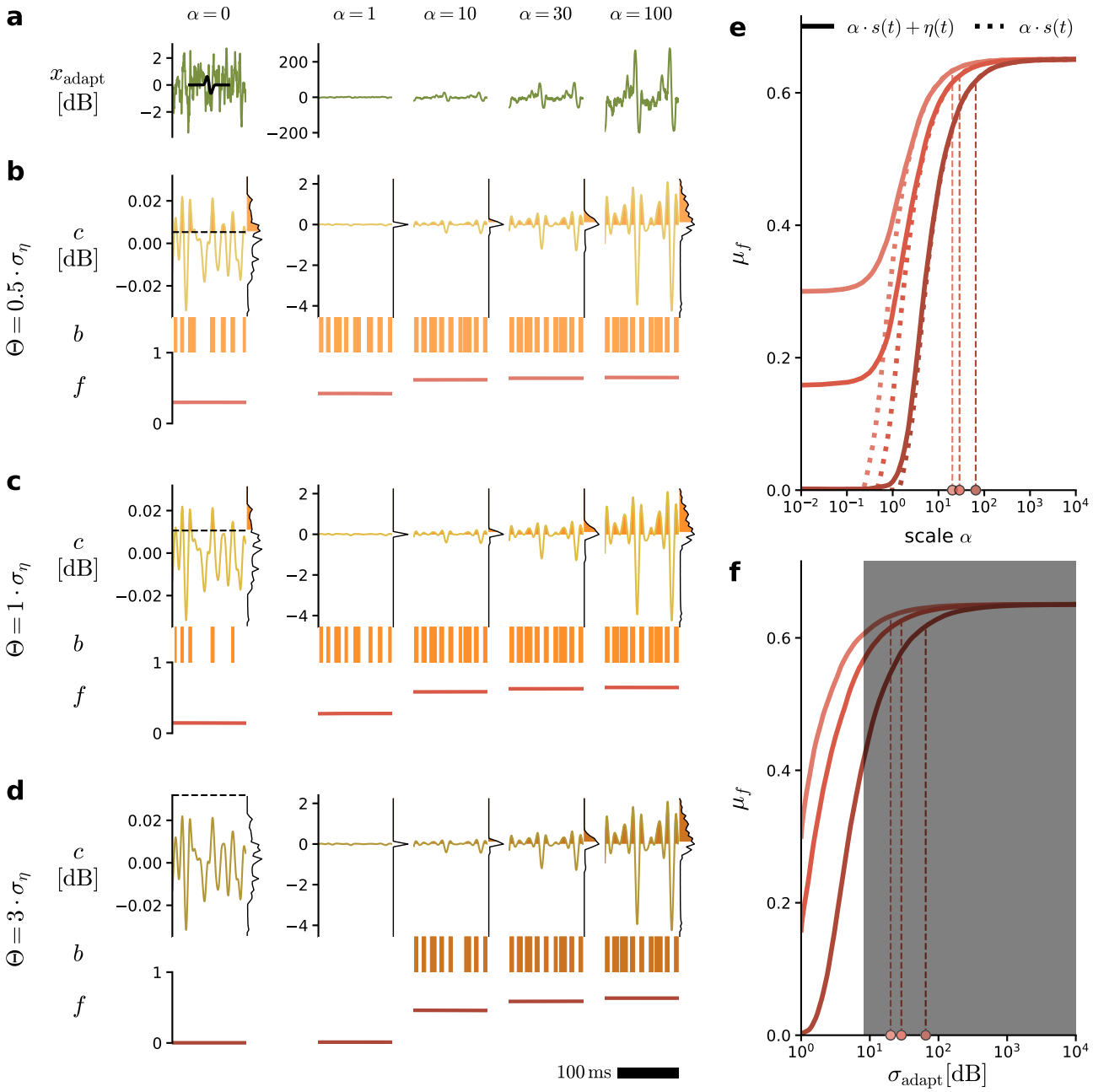

Fig. 6: Intensity invariance through thresholding and temporal averaging is mediated by the interaction of threshold value and noise floor. Input $x_{\text{adapt}}(t)$ consists of song component $s(t)$ scaled by α with optional noise component $\eta(t)$ and is transformed into single kernel response $c(t)$, binary response $b(t)$, and feature $f(t)$. Different color shades indicate different threshold values Θ (multiples of reference standard deviation σ_η of $c(t)$ for input $x_{\text{adapt}}(t) = \eta(t)$, with darker colors for higher Θ). **Left**: Noisy case: Example representations of $x_{\text{adapt}}(t)$ as well as $c(t)$, $b(t)$, and $f(t)$ for different α. **a**: $x_{\text{adapt}}(t)$ with kernel $k(t)$ in black. **b - d**: $c(t)$, $b(t)$, and $f(t)$ based on the same $x_{\text{adapt}}(t)$ from **a** but with different Θ. **Right**: Average value μ_f of $f(t)$ for each Θ from **b - d**. Dots indicate 95 % curve span (noisy case). **e**: μ_f over a range of α, once for the noisy case (solid lines) and once for the noiseless case (dotted lines). **f**: Noisy case: μ_f over the standard deviation of input x_{adapt} corresponding to the values of α shown in **e**. Shaded area indicates standard deviations that would be capped in the output $x_{\text{adapt}}(t)$ of the previous transformation pair (see Fig. 5cd).

3.2 Intensity invariance of species-specific feature representations

Having established both the meaning of the feature value and the mechanism of intensity invariance by thresholding and temporal averaging, the question remains how this mechanism acts on a set of features $f_i(t)$ based on different species-specific songs (Fig. 7a). The previous analysis was repeated with three different kernels $k_i(t)$ using a single kernel-specific threshold value Θ_i; and the resulting average feature values μ_{f_i} were plotted over α (Fig. 7bc). Additionally, 2D feature spaces spanned by each pair of $f_i(t)$ were plotted to investigate the separability of species-specific songs based on the feature representation in dependence of α (Fig. 7de). Each species-specific combination of μ_{f_i} follows a trajectory through feature space that develops with α. These trajectories correspond to the transient regime between the constant (noise) regime and the saturation regime, which are only visible as the start and end points of the trajectories, respectively. The horizontal dashes in the colorbars indicate the range of α that corresponds to the transient regime across $f_i(t)$ for each species.

In the noiseless case, each μ_{f_i} is 0 for small α across all species (Fig. 7b) because $c_i(t)$ never exceeds Θ_i. Accordingly, each trajectory starts at the origin of the feature space (Fig. 7d). For larger α, all μ_{f_i} saturate at individual values whose combination differs between species, so that the songs of each species are eventually represented by distinct points in feature space. However, the species-specific trajectories cross each other at numerous points, which means that the songs of two species — each at a specific α — can result in the same combination of μ_{f_i}. Furthermore, the specific value of α at which μ_{f_i} saturates depends on $f_i(t)$ and the species: For *C. mollis*, all μ_{f_i} saturate around the same α, while *O. rufipes* exhibits considerable variation between the three $f_i(t)$. The larger the variation in saturation points between $f_i(t)$, the stronger the curvature of the trajectory through feature space.

In the noisy case, μ_{f_i} is non-zero even for the smallest α (Fig. 7c) because the addition of the noise component $\eta(t)$ to input $x_{\text{adapt}}(t)$ drives $c_i(t)$ above Θ_i regardless of the song component $s(t)$. The starting value of μ_{f_i} is the same across all $f_i(t)$ and species by construction of the specific Θ_i. In consequence, the trajectories through feature space do not start at the origin but rather at approximately the same point along the diagonal (Fig. 7e). For larger α, all μ_{f_i} saturate at the same values as in the noiseless case, as expected from the previous analysis (Fig. 6e). However, the trajectories now move a much shorter distance through feature space for a similar range of α due to the lower SNR of $f_i(t)$ between noise regime and saturation regime, which increases the likelihood of trajectories crossing each other. Finally, the values of α at which μ_{f_i} saturate for a given species are slightly higher in the noisy case, but the variation between $f_i(t)$ remains largely unchanged.

In summary, even a comparably small set of three features $f_i(t)$ can, in principle, represent

different species-specific songs at distinct points in feature space, regardless of the presence of noise. However, this only holds for sufficiently large α that allow $f_i(t)$ to reach a saturation regime. During the transient regime, the species-specific combination of μ_{f_i} can very well be the same for two or more different species at specific α, although this may be alleviated by the inclusion of additional $f_i(t)$. Overall, the results of this analysis suggest that Θ_i should rather be choosen in favor of a higher SNR (Θ_i just above pure-noise $c_i(t)$) than a lower saturation point ($\Theta_i \to 0$). First, because this reduces the density of trajectories through feature space, and second, because the capping of $x_{\text{adapt}}(t)$ by the previous transformation pair likely renders the saturation point of $f_i(t)$ less relevant.

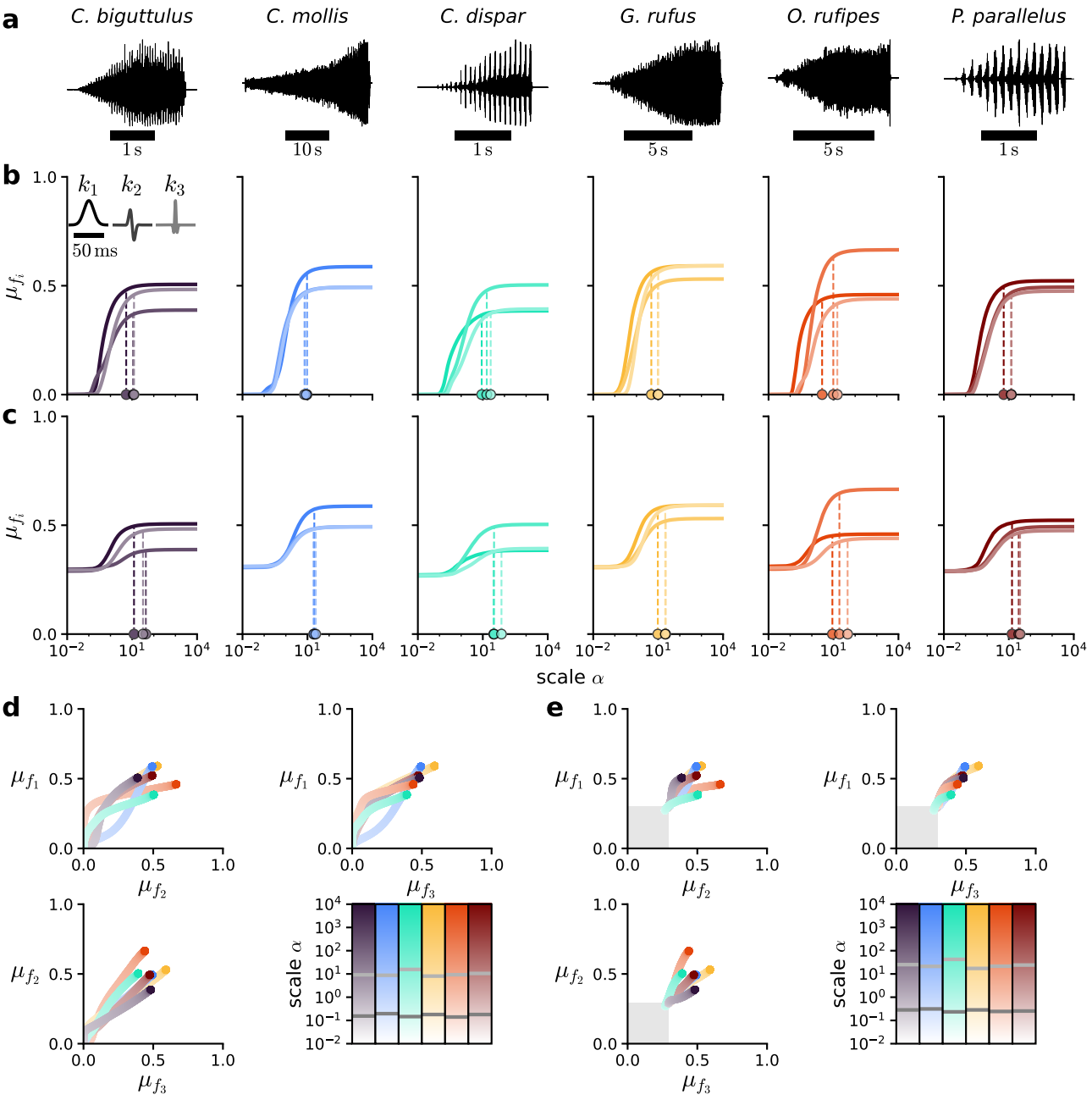

Fig. 7: Feature representation of different species-specific songs saturates at different points in feature space. Same input and processing as in Fig. 6 but with three different kernels k_i, each with a single kernel-specific threshold value $\Theta_i = 0.5 \cdot \sigma_{\eta_i}$. **a**: Examples of species-specific grasshopper songs. **Middle**: Average value μ_{f_i} of each feature $f_i(t)$ over α per species (averaged over songs and recordings, see appendix Figs. 16 and 17). Different color shades indicate different kernels k_i. Dots indicate 95 % curve span per k_i. **b**: Noiseless case. **c**: Noisy case. **Bottom**: 2D feature spaces spanned by each pair of $f_i(t)$. Each trajectory corresponds to a species-specific combination of μ_{f_i} that develops with α (colorbars). Horizontal dashes in the colorbar indicate 5 % (dark grey) and 95 % (light grey) curve span of the norm across all three μ_{f_i} per species. **d**: Noiseless case. **e**: Noisy case. Shaded areas indicate the average minimum μ_{f_i} across all species-specific trajectories.

3.3 Intensity invariance along the full model pathway

Through the previous analyses, we could establish two mechanisms of intensity invariance: Logarithmic compression and adaptation as well as thresholding and temporal averaging. While each transformation pair by itself can provide some level of invariance, certain results suggest that the first mechanism may actually limit or even nullify the effect of the second mechanism. In the following sections, we investigate the combined effect of both mechanisms along the full model pathway (Fig. 8) and explore the consequences of disabling the first mechanism by skipping the logarithmic compression step (Fig. 9).

3.3.1 Including logarithmic compression

For this analysis, input $x_{\mathrm{raw}}(t)$ — including both song component $s(t)$ and noise component $\eta(t)$ — was rescaled and processed throughout all steps of the model pathway (Fig. 8a) up to the feature set $f_i(t)$. As before, the standard deviation was used as intensity metric for each resulting representation except $b_i(t)$ and $f_i(t)$. For $f_i(t)$, the average feature value μ_{f_i} was used, while $b_i(t)$ was omitted from the analysis. Plotting each intensity metric over α (Fig. 8b) reinforces many of the previous observations. For ease of visualization, the kernel-specific curves for $c_i(t)$ and $f_i(t)$ were summarized by their median. Representations prior to logarithmic compression — $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$ — show a linear increase of the intensity metric for larger α on a double-logarithmic scale. Representations after logarithmic compression — $x_{\log}(t)$, $x_{\mathrm{adapt}}(t)$, and $c_i(t)$ — are the first to reach a saturation regime and do so at approximately the same α because they are separated only by linear transformations. Feature set $f_i(t)$ reaches a saturation regime, as well. But contrary to previous results, the saturation point of $f_i(t)$ appears below that of $c_i(t)$, which suggests that the second mechanism of thresholding and temporal averaging can indeed improve intensity invariance beyond the first mechanism of logarithmic compression and adaptation. The difference in saturation points is best illustrated based on the ratio of each intensity metric to the respective pure-noise reference value (Fig. 8d). However, compressing $f_i(t)$ into a median across $k_i(t)$ conceils many kernel-specific details. It is therefore necessary to consider the development of each $f_i(t)$ over α separately (Fig. 8c). Indeed, all 40 $f_i(t)$ in the set reach a saturation regime for sufficiently large α. The saturated μ_{f_i} are distributed over a range of values — which is the prerequisite for forming species-specific combinations — but are limited to a rather small subset of possible values between 0 and 1. Based on previous results (Fig. 6f), this is likely due to the capping of $x_{\mathrm{adapt}}(t)$ that prevents $f_i(t)$ from reaching its intrinsic saturation value; but this cannot be confirmed until the following analysis (Fig. 9). Looking at the kernel-specific SNR values of $c_i(t)$ over α (Fig. 8e) and $f_i(t)$ over α (Fig. 8f) reveals a high degree of variation between different $k_i(t)$. Certain $f_i(t)$ achieve much higher SNR values than $c_i(t)$ for the same α due to the former's capacity for arbitrarily

low pure-noise responses ($\mu_{f_i} \to 0$) and hence arbitrarily high SNR values. Finally, the question remains whether the suspected improvement of intensity invariance by $f_i(t)$ beyond $c_i(t)$ holds at the level of individual $k_i(t)$. The single saturation points based on the median across $k_i(t)$ for $c_i(t)$ and $f_i(t)$ are expanded into distributions of kernel-specific saturation points (Fig. 8g). For $c_i(t)$, the distribution is rather narrow and corresponds well to the single saturation point based on the median. For $f_i(t)$, however, the distribution is much broader and is not centered around the single saturation point based on the median but rather shifted towards lower α. Care must be taken when interpreting the height of either distribution due to the logarithmic scaling of the underlying α axis. Nevertheless, the overall pattern suggests that specific $f_i(t)$ can reach a saturation regime at lower α than their $c_i(t)$ counterparts. Therefore, the effect of thresholding and temporal averaging on intensity invariance is not necessarily nullified by the previous logarithmic compression and adaptation, which means that both mechanisms can, in principle, work together towards an intensity-invariant song representation.

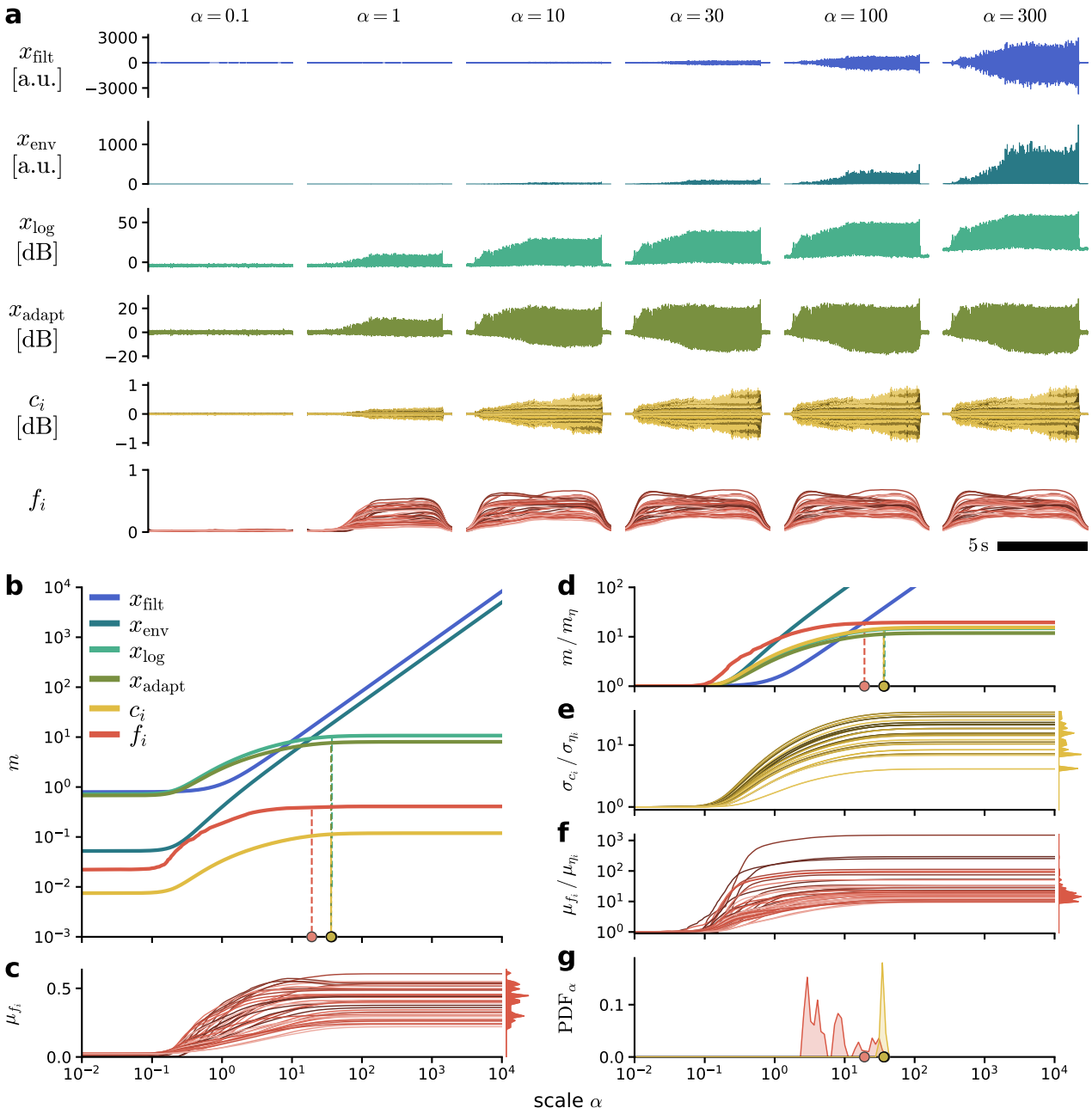

Fig. 8: Step-wise emergence of intensity-invariant song representations along the model pathway. Input $x_{\text{raw}}(t)$ consists of song component $s(t)$ scaled by α with added noise component $\eta(t)$ and is processed up to the feature set $f_i(t)$. Different color shades indicate different types of Gabor kernels with specific lobe number n and either + or − sign, sorted (dark to light) first by increasing n and then by sign ($1 \leq n \leq 4$; first +, then − for each n; five kernel widths σ of 1, 2, 4, 8, and 16 ms per type; 8 types, 40 kernels in total). **a**: Example representations of $x_{\text{filt}}(t)$, $x_{\text{env}}(t)$, $x_{\text{log}}(t)$, $x_{\text{adapt}}(t)$, $c_i(t)$, and $f_i(t)$ for different α. **b**: Intensity metrics over α. For $c_i(t)$ and $f_i(t)$, the median over kernels is shown. Dots indicate 95 % curve span for $x_{\text{log}}(t)$, $x_{\text{adapt}}(t)$, $c_i(t)$, and $f_i(t)$. **c**: Average value μ_{f_i} of each feature $f_i(t)$ over α. **d**: Ratios of intensity metrics to the respective reference value for input $x_{\text{raw}}(t) = \eta(t)$. For $c_i(t)$ and $f_i(t)$, the median over kernel-specific ratios is shown. **e**: Ratios of standard deviation σ_{c_i} of each $c_i(t)$. **f**: Ratios of μ_{f_i}. **g**: Distributions of kernel-specific α that correspond to 95 % curve span for $c_i(t)$ and $f_i(t)$. Dots indicate the values from **b**.

3.3.2 Excluding logarithmic compression

The previous analysis was repeated in exactly the same way as before, except that the logarithmic compression of $x_{\mathrm{env}}(t)$, Eq. 3, was skipped in order to disable the first mechanism of intensity invariance. Consequently, $x_{\mathrm{adapt}}(t)$ is merely a highpass filtered version of $x_{\mathrm{env}}(t)$; and $x_{\log}(t)$ is missing entirely (Fig. 9a). As expected, all representations prior to the thresholding nonlinearity $H(c_i - \Theta_i)$ — $x_{\mathrm{filt}}(t)$, $x_{\mathrm{env}}(t)$, $x_{\mathrm{adapt}}(t)$, and $c_i(t)$ — show a linear increase of the intensity metric for larger α, while $f_i(t)$ is the only representation to reach a saturation regime (Fig. 9bd). The saturated μ_{f_i} are distributed over a much broader range of values than in the previous analysis (Fig. 9c). Intriguingly, the distribution of μ_{f_i} is symmetric around a value of 0.5. This is relevant because every kernel $k^+(t)$ in the underlying kernel set has a counterpart of opposite sign that is otherwise identical, so that $k^+(t) = -k^-(t)$. The responses of $k^+(t)$ and $k^-(t)$ to the same input $x_{\mathrm{adapt}}(t)$ are also inverted because convolution is a linear operation: $c^+(t) = -c^-(t)$. The distributions of $c^+(t)$ and $c^-(t)$ are hence inverted to each other, as well: $p(c^+) = p(-c^-)$. Based on Eq. 23, transforming $c^+(t)$ and $c^-(t)$ further using the same Θ thus results in two complementary features $f^+(t)$ and $f^-(t)$ that are symmetric around 0.5, so that $f^+(t) = 1 - f^-(t)$. Of course, this symmetry throughout the feature representation goes hand in hand with a substantial degree of redundancy and is hardly expected to be present in the actual grasshopper auditory system. But the fact that the saturated μ_{f_i} are distributed symmetrically around 0.5 provides concrete evidence that each $f_i(t)$ is able to reach its intrinsic saturation value in the absence of logarithmic compression (Fig. 9c), which is otherwise prevented by the capping of $x_{\mathrm{adapt}}(t)$, as seen during previous analyses (Fig. 6f and Fig. 8c). Otherwise, there appear to be no major differences in the development of $f_i(t)$ over α compared to the previous analysis, neither on the kernel-specific SNR values (Fig. 9e) nor on the distribution of kernel-specific saturation points (Fig. 9f). Overall, the most substantial consequence of skipping the logarithmic compression is that it allows $f_i(t)$ to reach its intrinsic saturation value. If this results in a wider range of μ_{f_i} across the feature set, it should be beneficial for forming species-specific combinations. However, this depends on multiple different factors such as the choice of $k_i(t)$ and Θ_i as well as the structure and distribution of the specific song and is hence not guaranteed simply by disabling logarithmic compression.

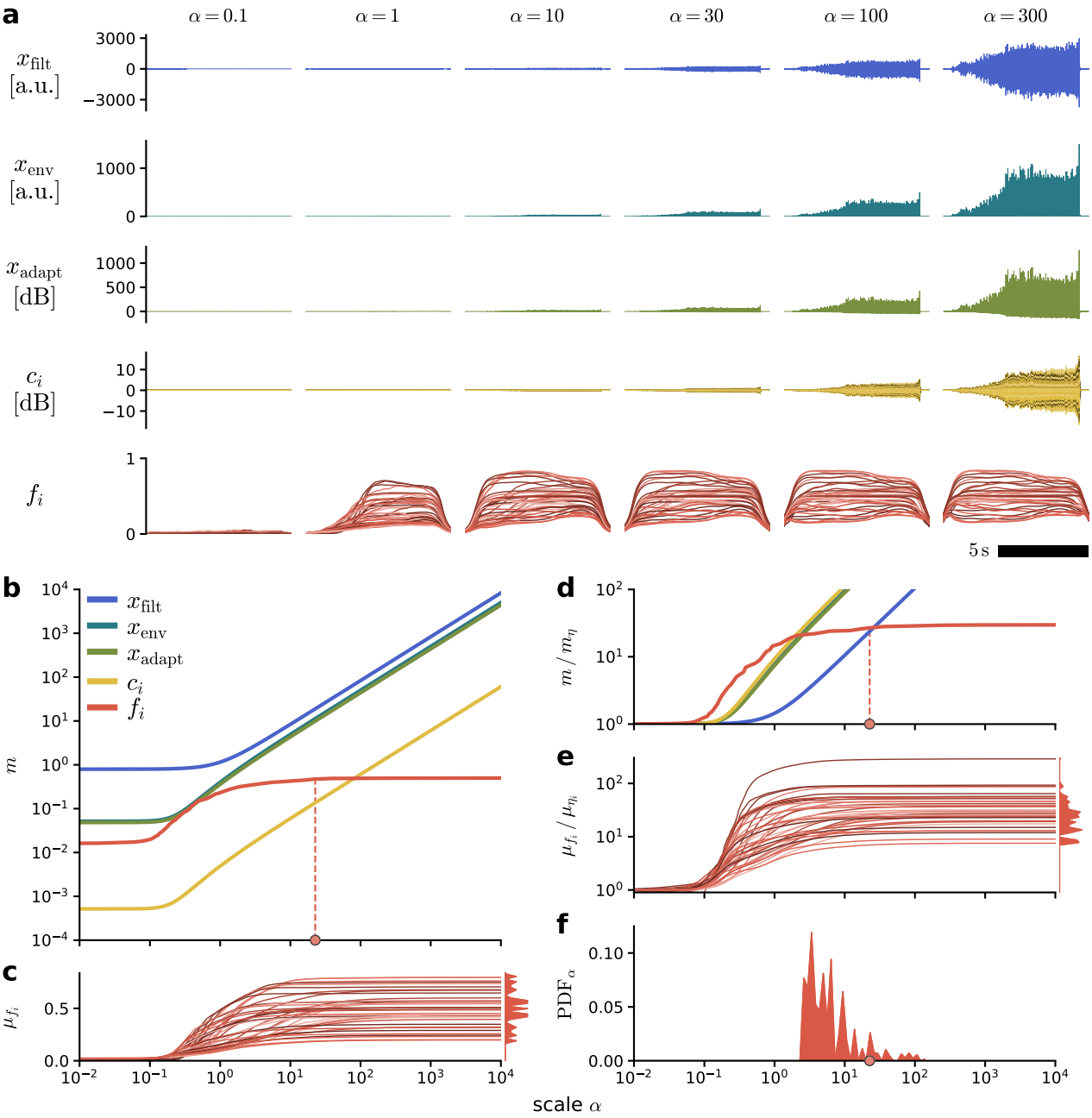

Fig. 9: Effects of disabling logarithmic compression on intensity invariance along the model pathway. Input $x_{\text{raw}}(t)$ consists of song component $s(t)$ scaled by α with added noise component $\eta(t)$ and is processed up to the feature set $f_i(t)$, skipping $x_{\log}(t)$. Different color shades indicate different types of Gabor kernels with specific lobe number n and either + or − sign, sorted (dark to light) first by increasing n and then by sign ($1 \leq n \leq 4$; first +, then − for each n; five kernel widths σ of 1, 2, 4, 8, and 16 ms per type; 8 types, 40 kernels in total). **a**: Example representations of $x_{\text{filt}}(t)$, $x_{\text{env}}(t)$, $x_{\text{adapt}}(t)$, $c_i(t)$, and $f_i(t)$ for different α. **b**: Intensity metrics over α. For $c_i(t)$ and $f_i(t)$, the median over kernels is shown. Dots indicate 95 % curve span for $f_i(t)$. **c**: Average value μ_{f_i} of each feature $f_i(t)$ over α. **d**: Ratios of intensity metrics to the respective reference value for input $x_{\text{raw}}(t) = \eta(t)$. For $c_i(t)$ and $f_i(t)$, the median over kernel-specific ratios is shown. **e**: Ratios of μ_{f_i}. **f**: Distribution of kernel-specific α that correspond to 95 % curve span for $f_i(t)$. Dots indicate the value from **b**.

3.3.3 Intensity invariance in a naturalistic setting

So far, the analyses on intensity invariance were based on synthetically generated input signals, since these allow for a systematic manipulation of the mixture of song component $s(t)$ and noise component $\eta(t)$ over an arbitrary range of scales α. Now, the question remains how the model pathway performs under more naturalistic conditions. The previous analysis of the full model pathway (Fig. 8) was hence repeated, using field recordings of a song of *P. parallelus* as input $x_{\mathrm{raw}}(t)$ and a segment of background noise from the same recordings as pure-noise reference. Recordings were taken simultaneously at eight different distances d from the sender, ranging from 10 cm to 220 cm with intervals of 30 cm between microphones. The precise value of α that corresponds to a given d cannot be determined in a straightforward manner, but α is expected to be inversely proportional to d based on the inverse-square law of sound propagation. All intensity metrics and ratios thereof were hence plotted over $1/d$ on a double-logarithmic scale, which is insofar comparable to previous analyses that a decade on the $1/d$ axis corresponds to a decade on the α axis. To complicate matters further, the $1/d$ axis is sampled too sparsely to determine saturation points as before based on the 95 % curve span. Instead, one has to rely on the slope of the curve to assess if, and at which $1/d$, a given representation reaches a saturation regime. Bearing these limitations in mind, the intensity metrics of each representation over $1/d$ (Fig. 10b) follow a pattern that is consistent with the results of the previous simulation-based analysis (Fig. 8b): The standard deviations of $x_{\mathrm{filt}}(t)$ and $x_{\mathrm{env}}(t)$ increase linearly with $1/d$, respectively. The standard deviations of $x_{\log}(t)$, $x_{\mathrm{adapt}}(t)$, and $c_i(t)$ show a weaker increase with $1/d$ and appear to approach, but not reach, a saturation regime for larger $1/d$. The average feature values μ_{f_i} of $f_i(t)$ show an even weaker increase with $1/d$ and appear to reach a saturation regime for $d = 40\,\mathrm{cm}$ and $d = 10\,\mathrm{cm}$, which is consistent across most $f_i(t)$ in the set (Fig. 10c). Saturation of $f_i(t)$ without saturation of $c_i(t)$ suggests that the input $x_{\mathrm{raw}}(t)$ at the smallest $d = 10\,\mathrm{cm}$ corresponds to a value of α between 10 and 20 based on comparison with the simulation-based analysis (Fig. 8b). The saturated μ_{f_i} are distributed over a comparably narrow range of values, which could in parts be a property of the songs of *P. parallelus* (see also Fig. 7bc). The ratios of each intensity metric to the respective pure-noise reference value are not aligned across representations (Fig. 10d) or kernels (Fig. 10ef) but serve to consolidate the previous observation that only $f_i(t)$ exhibits some degree of intensity invariance within the available range of $1/d$. Based on the current results, this intensity invariance of $f_i(t)$ in the field holds up to a distance of around 40 cm from the sender, decays steadily between 40 cm and 130 cm, and is substantially dimished for larger distances (Fig. 10a, bottom row).

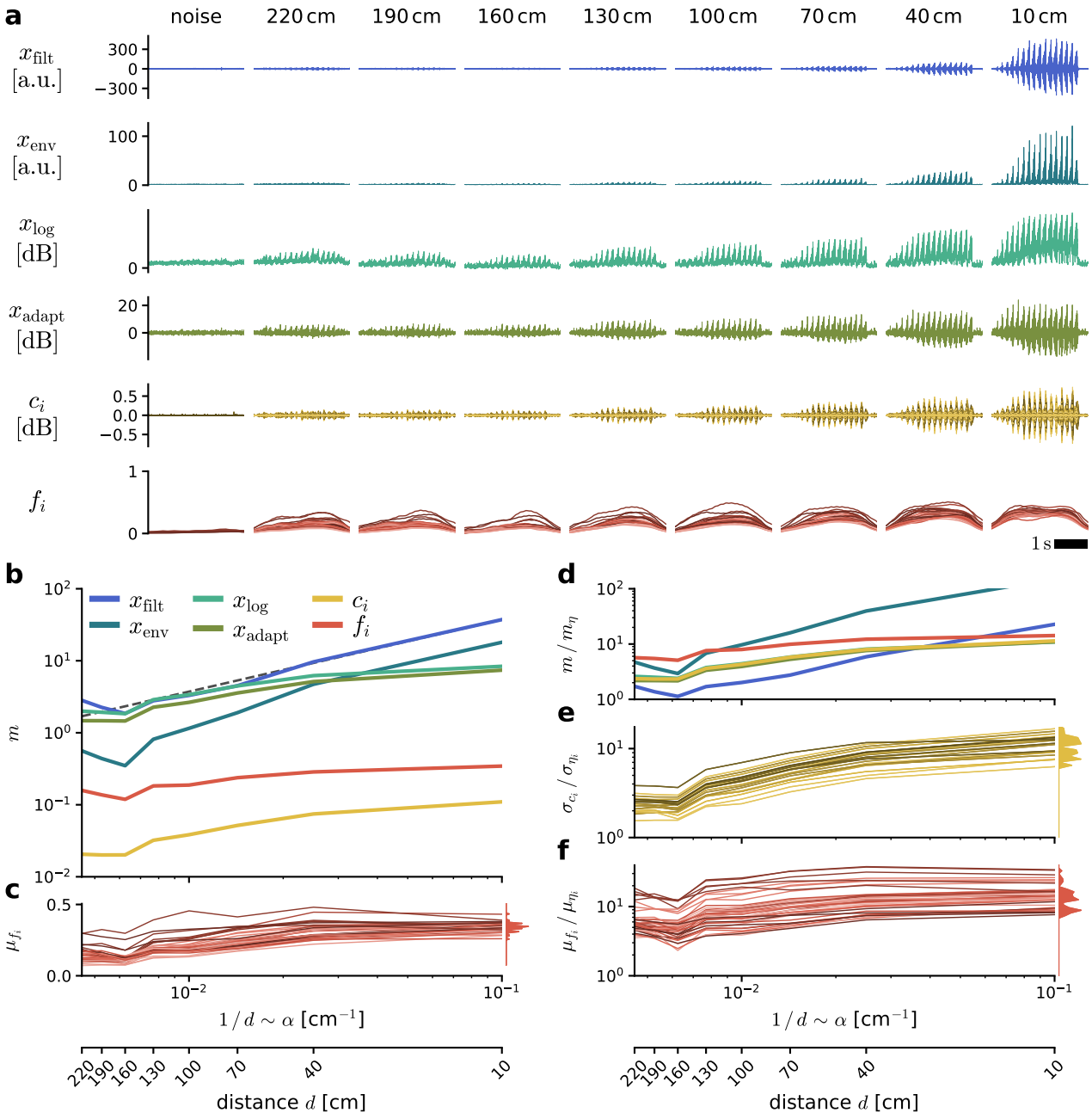

Fig. 10: Intensity invariance along the model pathway in a naturalistic setting. Input $x_{\mathrm{raw}}(t)$ consists of a song of *P. parallelus* recorded in the field at eight different distances d and is processed up to the feature set $f_i(t)$. Different color shades indicate different types of Gabor kernels with specific lobe number n and either + or − sign, sorted (dark to light) first by increasing n and then by sign ($1 \leq n \leq 4$; first +, then − for each n; five kernel widths σ of 1, 2, 4, 8, and 16 ms per type; 8 types, 40 kernels in total). **a**: $x_{\mathrm{filt}}(t)$, $x_{\mathrm{env}}(t)$, $x_{\mathrm{log}}(t)$, $x_{\mathrm{adapt}}(t)$, $c_i(t)$, and $f_i(t)$ at each d. A noise segment from the same recording is shown for reference. **b**: Intensity metrics over d. For $c_i(t)$ and $f_i(t)$, the median over kernels is shown. **c**: Average value μ_{f_i} of each feature $f_i(t)$ over d. **d**: Ratios of intensity metrics to the respective value obtained from the noise reference. For $c_i(t)$ and $f_i(t)$, the median over kernel-specific ratios is shown. **e**: Ratios of standard deviation σ_{c_i} of each $c_i(t)$. **f**: Ratios of μ_{f_i}.

3.4 Interspecific and intraspecific feature variability

In the final analysis of the current study, we investigated the variability of songs in the feature representation between different species and within the same species (Fig. 11). Naturally, a feature representation that is both consistent across different songs of the same species and sufficiently different between songs of different species is a fundamental prerequisite for species-specific song recognition. The data used in this analysis corresponds to the saturated μ_{f_i} of each $f_i(t)$ from the previous analysis of the full model pathway (Fig. 8c), using different songs of *O. rufipes* for the intraspecific comparisons and single songs from a number of species for the interspecific comparisons (also shown in Fig. 7a). Accordingly, each song is represented by 40 values of μ_{f_i} based on the same set of $f_i(t)$. For each comparison, μ_{f_i} from one song was plotted against μ_{f_i} from the other song, so that each dot within a subplot corresponds to a single feature $f_i(t)$. For the intraspecific comparisons (Fig. 11, upper triangular), the pairs of μ_{f_i} are distributed closely around the diagonal, with a minimum correlation coefficient of $\rho = 0.85$, a maximum of $\rho = 0.99$, and a median of $\rho = 0.92$. A given $f_i(t)$ thus tends to have a similar μ_{f_i} across different songs of the same species. In contrast, the pairs of μ_{f_i} for the interspecific comparisons (Fig. 11, lower triangular) are distributed in a variety of different ways, most in broader clouds (e.g. *C. biguttulus* vs. *C. mollis*) but some more narrowly around the diagonal (e.g. *P. parallelus* vs. *C. dispar*). The correlation coefficients ρ vary widely between different interspecific comparisons, with a minimum of $\rho = -0.1$, a maximum of $\rho = 0.92$, and a median of $\rho = 0.53$. A given $f_i(t)$ therefore tends to have a less similar μ_{f_i} across different species than within the same species, although certain exeptions exist (Fig. 11, lower right). Accordingly, the feature representation that is generated by the model pathway is, in principle, suitable for the distinction between different species-specific songs. However, even the songs of the same species are subject to considerable variability in various aspects and depending on a multitude of external and internal factors, which cannot be fully captured based on a limited number of songs. The results of the current analysis are hence to be treated as a proof-of-concept that paves the way towards more comprehensive investigations on the details of song representation in feature space, including the effects of different parameters of the model pathway as well as the inclusion of additional songs and species to reflect the complexity of natural song variation.

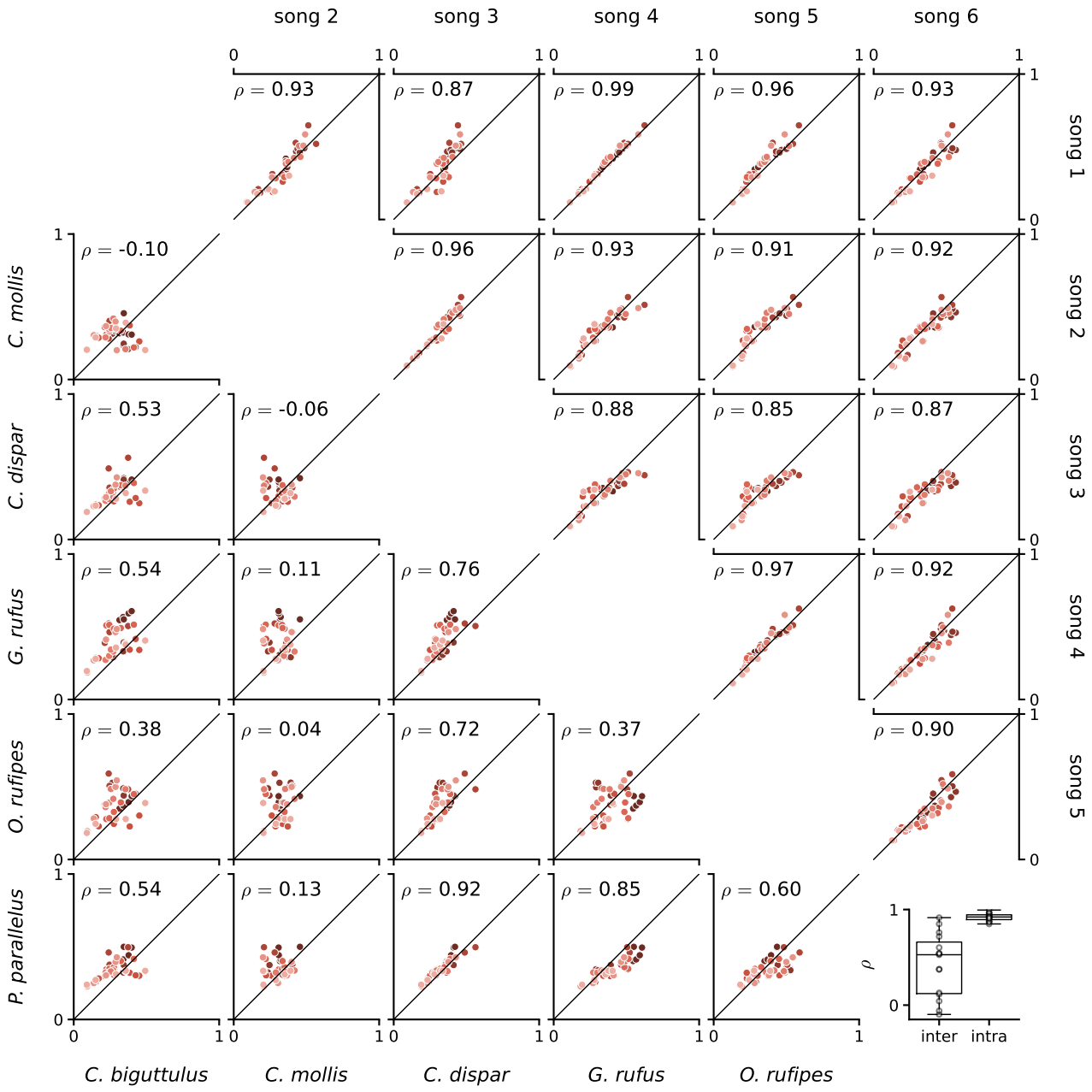

Fig. 11: Interspecific and intraspecific feature variability. Average value μ_{f_i} of each feature $f_i(t)$ against its counterpart from a 2nd feature set based on a different input $x_{\text{raw}}(t)$. Each dot within a subplot represents a single feature $f_i(t)$. Different color shades indicate different types of Gabor kernels with specific lobe number n and either + or − sign, sorted (dark to light) first by increasing n and then by sign ($1 \leq n \leq 4$; first +, then − for each n; five kernel widths σ of 1, 2, 4, 8, and 16 ms per type; 8 types, 40 kernels in total). Data is based on the analysis underlying Fig 8. **Lower triangular**: Interspecific comparisons between single songs of different species. **Upper triangular**: Intraspecific comparisons between different songs of a single species (*O. rufipes*). **Lower right**: Distribution of correlation coefficients ρ for each interspecific and intraspecific comparison. Dots indicate single ρ values.

4 Conclusions & outlook

Song recognition pathway: Grasshopper vs. model:
The model pathway includes a rather large number of Gabor kernels compared to the 15 to 20 ascending neurons in the grasshopper auditory system (Stumpner and Ronacher 1991).

Definition of invariance (general, systemic):
Invariance = Property of a system to maintain a stable output with respect to a set of relevant input parameters (variation to be represented) but irrespective of one or more other parameters (variation to be discarded) → Selective input-output decorrelation

Definition of intensity invariance (context of neurons and songs):
Intensity invariance = Time scale-selective sensitivity to certain faster amplitude dynamics (song waveform, small-scale AM) and simultaneous insensitivity to slower, more sustained amplitude dynamics (transient baseline, large-scale AM, current overall intensity level)
→ Without time scale selectivity, any fully intensity-invariant output will be a flat line

Log-HP: 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

- 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. **??**):
→ 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

Thresh-LP: Implication for intensity invariance:
- Role of song periodicity for feature representation!

- 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)

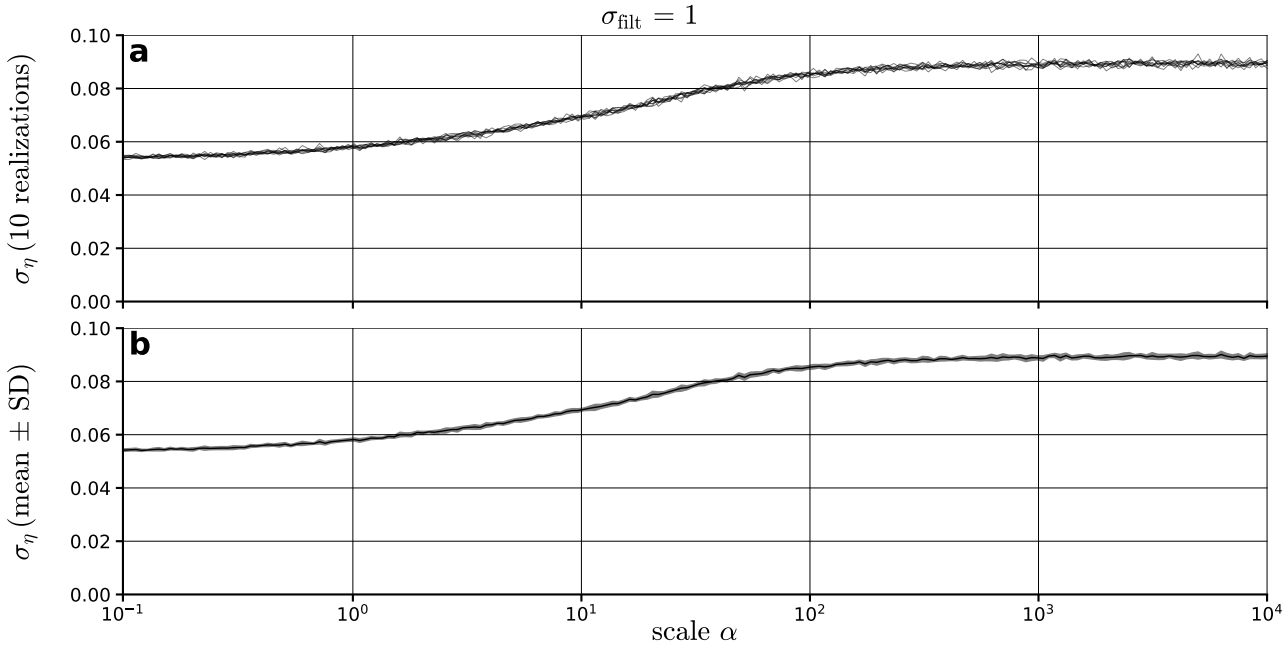

Fig. 12:

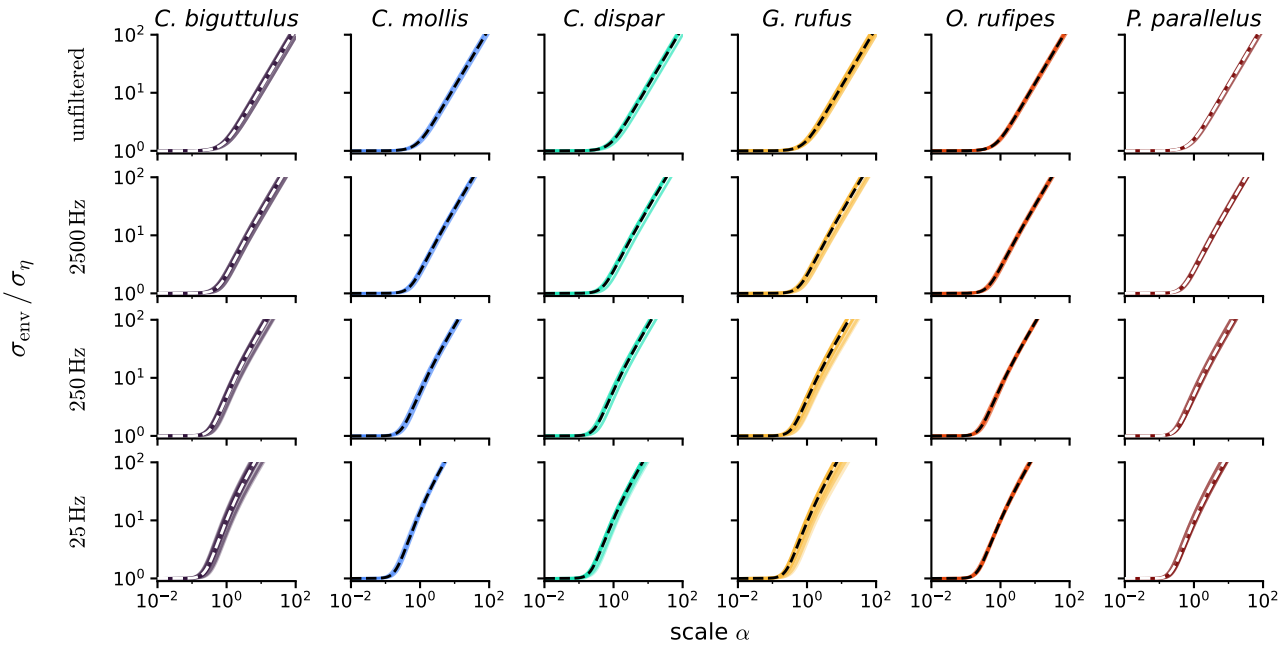
C. biguttulus
C. mollis
C. dispar
G. rufus
O. rufipes
P. parallelus
unfiltered
2500 Hz
250 Hz
25 Hz
$\sigma_{\mathrm{env}} / \sigma_\eta$
scale α

Fig. 13:

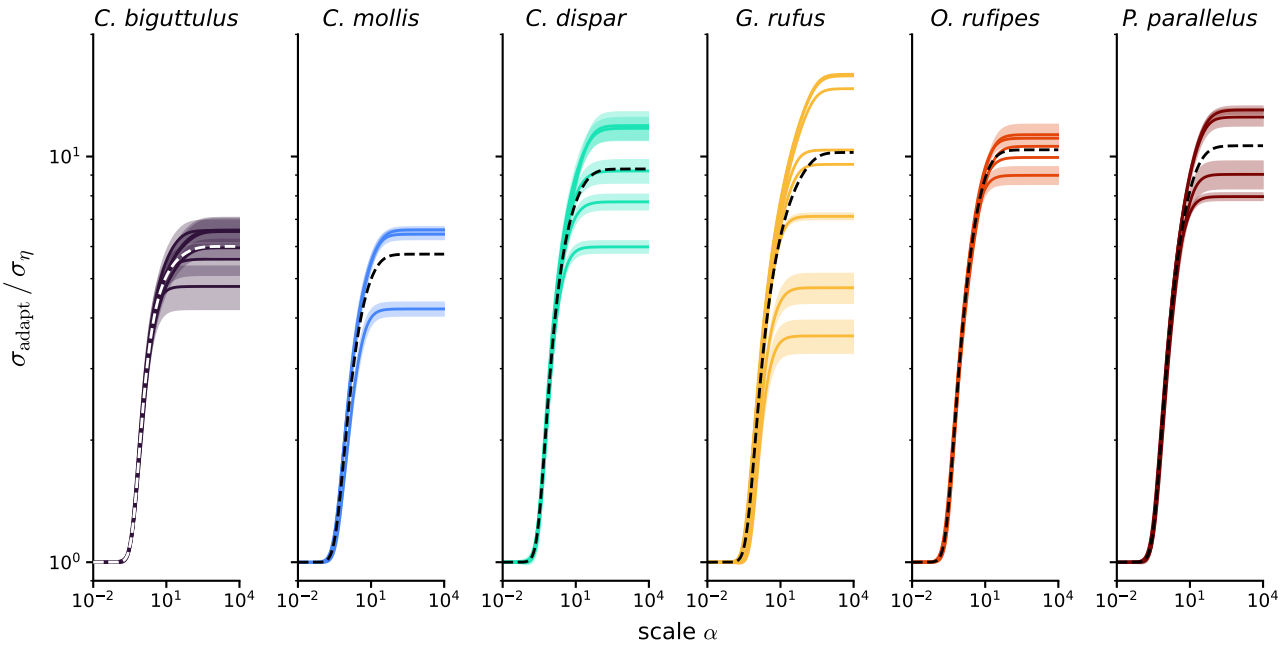
C. biguttulus
C. mollis
C. dispar
G. rufus
O. rufipes
P. parallelus
$\sigma_{\mathrm{adapt}} / \sigma_\eta$
scale α

Fig. 14:

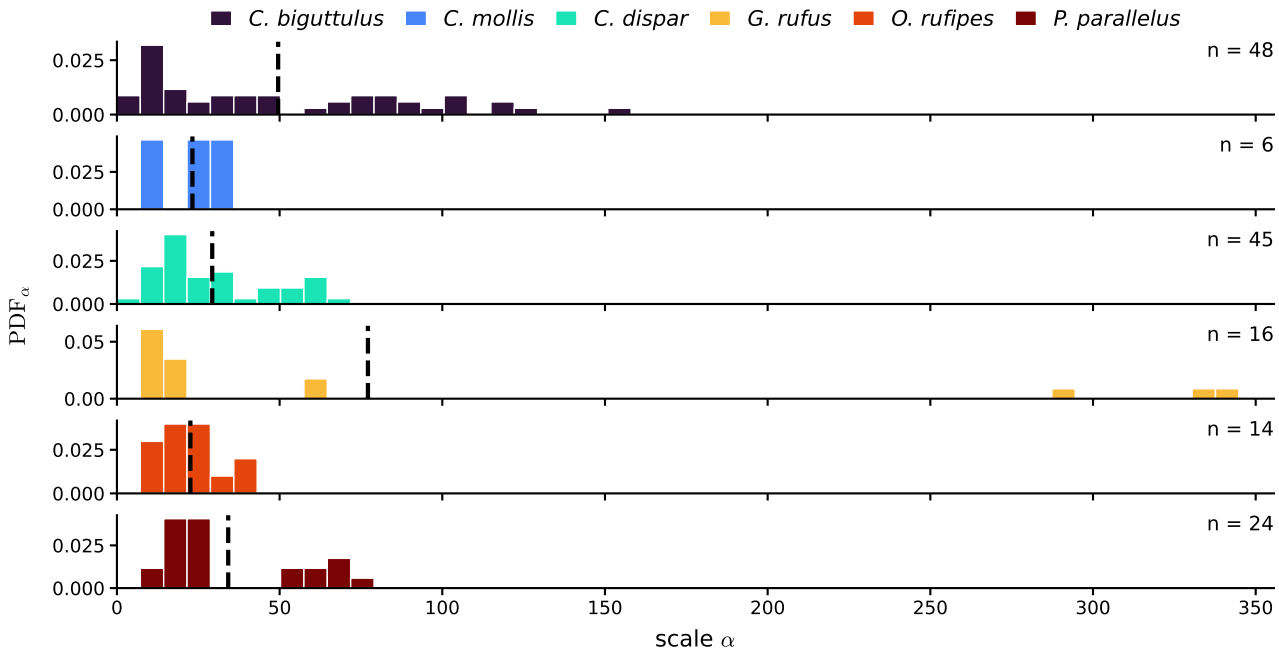
C. biguttulus
C. mollis
C. dispar
G. rufus
O. rufipes
P. parallelus
n = 48
n = 6
n = 45
n = 16
n = 14
n = 24
PDF$_\alpha$
scale α

Fig. 15:

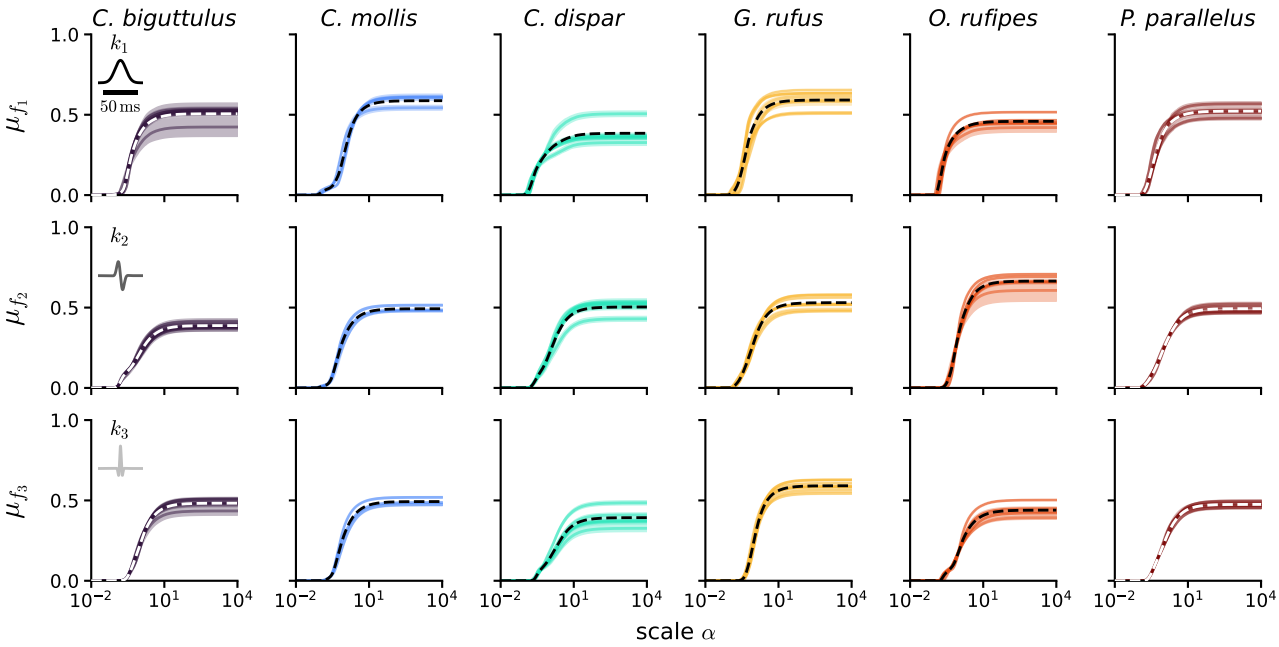
C. biguttulus
C. mollis
C. dispar
G. rufus
O. rufipes
P. parallelus
k_1
50 ms
k_2
k_3
μ_{f_1}
μ_{f_2}
μ_{f_3}
scale α

Fig. 16:

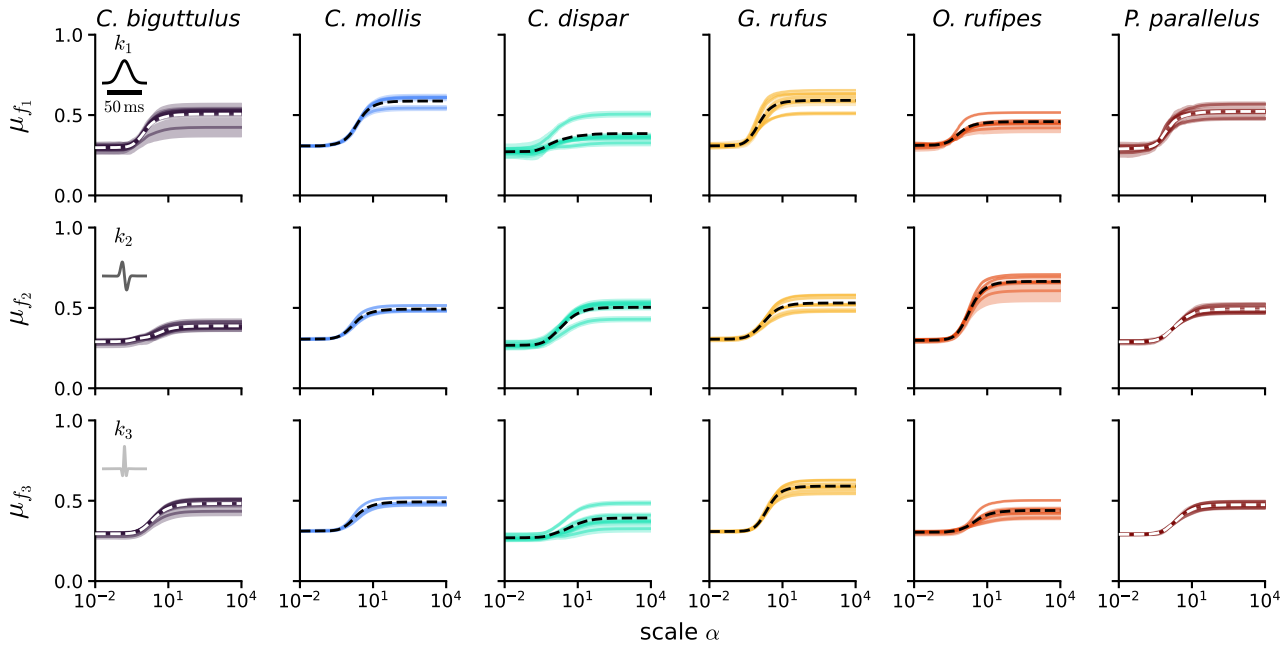
C. biguttulus
C. mollis
C. dispar
G. rufus
O. rufipes
P. parallelus
k_1
50 ms
k_2
k_3
μ_{f_1}
μ_{f_2}
μ_{f_3}
1.0
0.5
0.0
10^{-2}
10^{1}
10^{4}
scale α

Fig. 17:

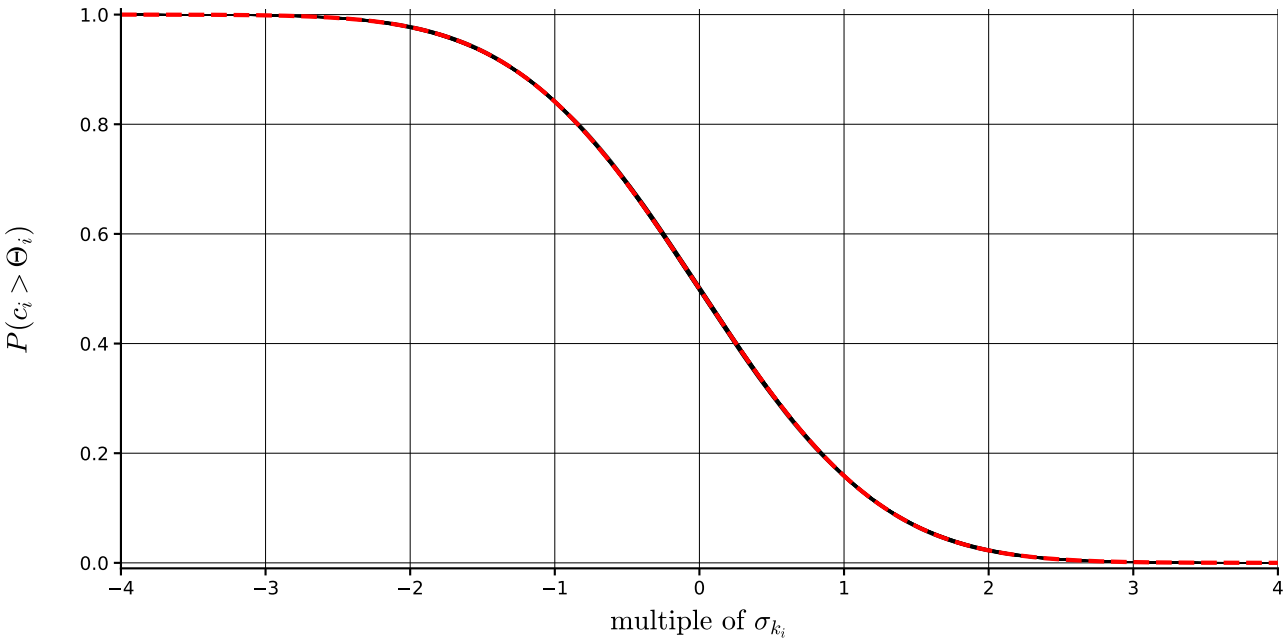
$P(c_i > \Theta_i)$
1.0
0.8
0.6
0.4
0.2
0.0
−4
−3
−2
−1
0
1
2
3
4
multiple of σ_{k_i}

Fig. 18:

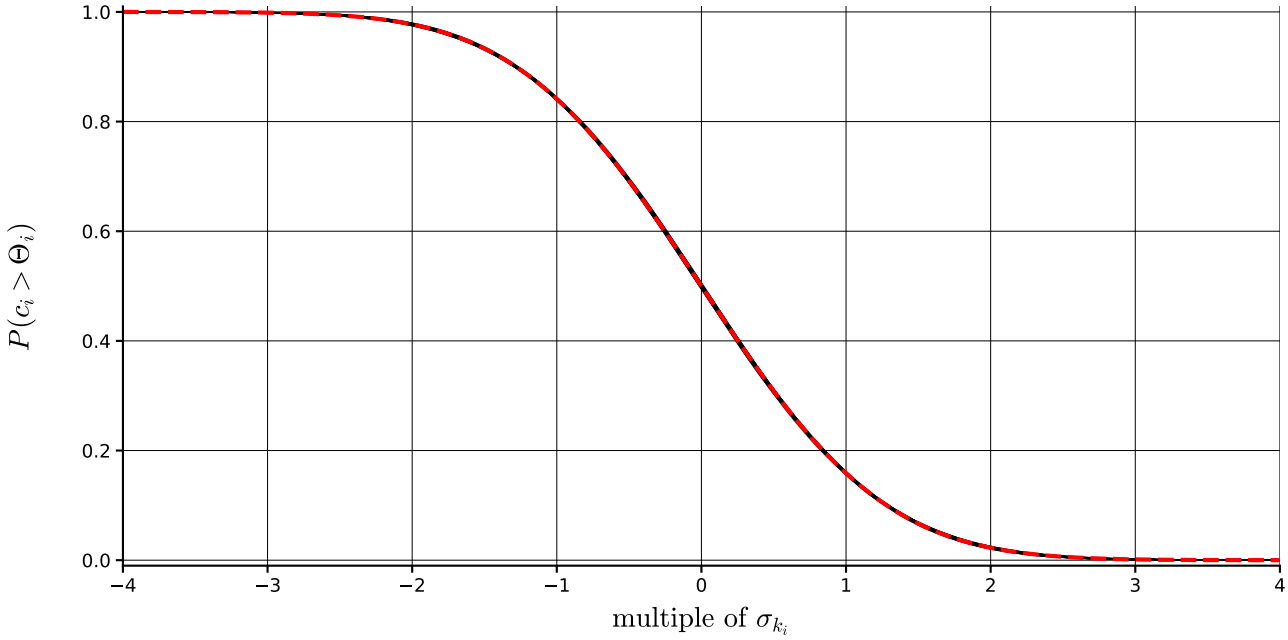
1.0
0.8
0.6
0.4
0.2
0.0
$P(c_i > \Theta_i)$
−4
−3
−2
−1
0
1
2
3
4
multiple of σ_{k_i}

Fig. 19:

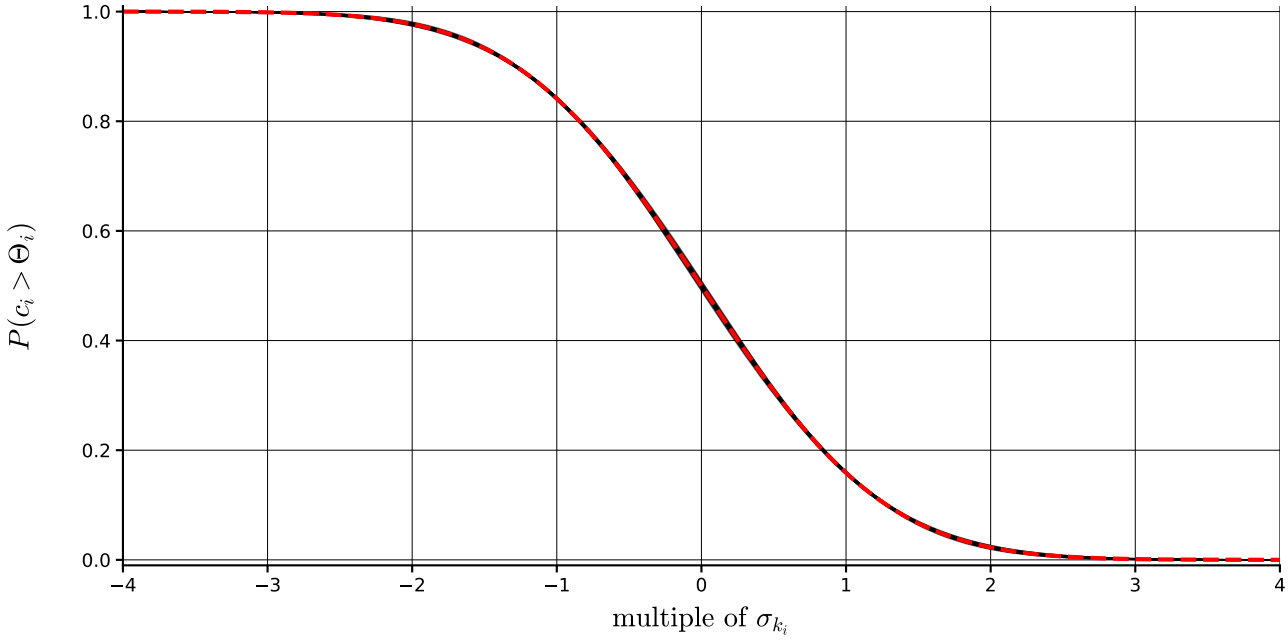
1.0
0.8
0.6
0.4
0.2
0.0
$P(c_i > \Theta_i)$
−4
−3
−2
−1
0
1
2
3
4
multiple of σ_{k_i}

Fig. 20:

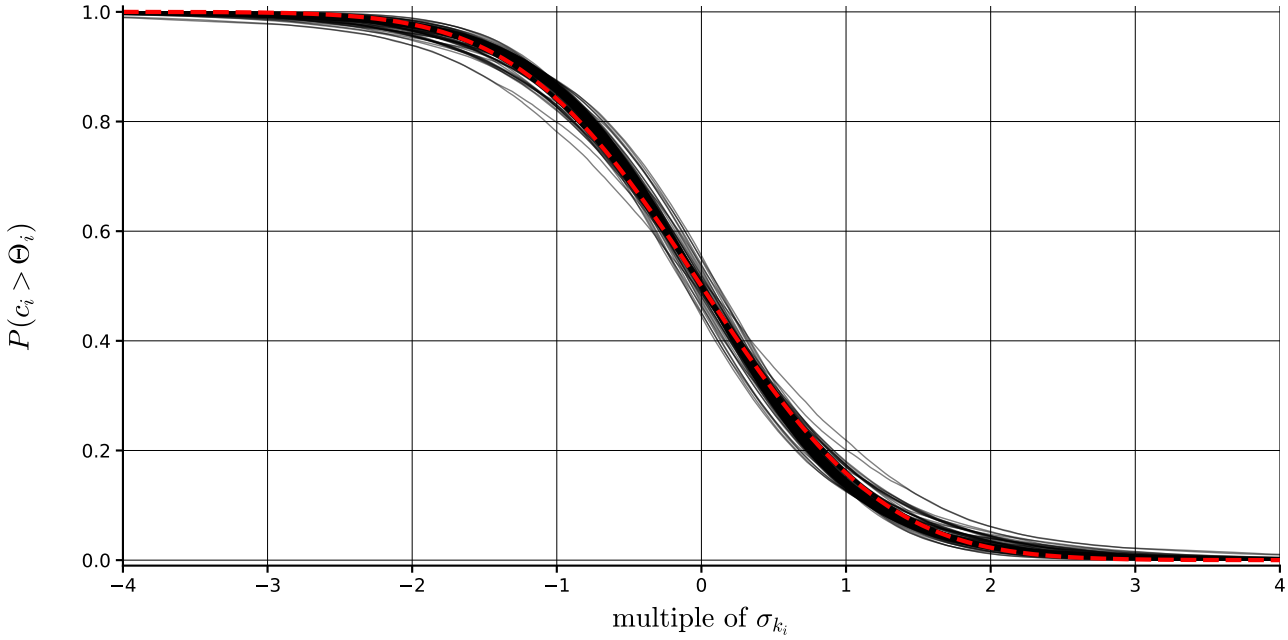

Fig. 21:

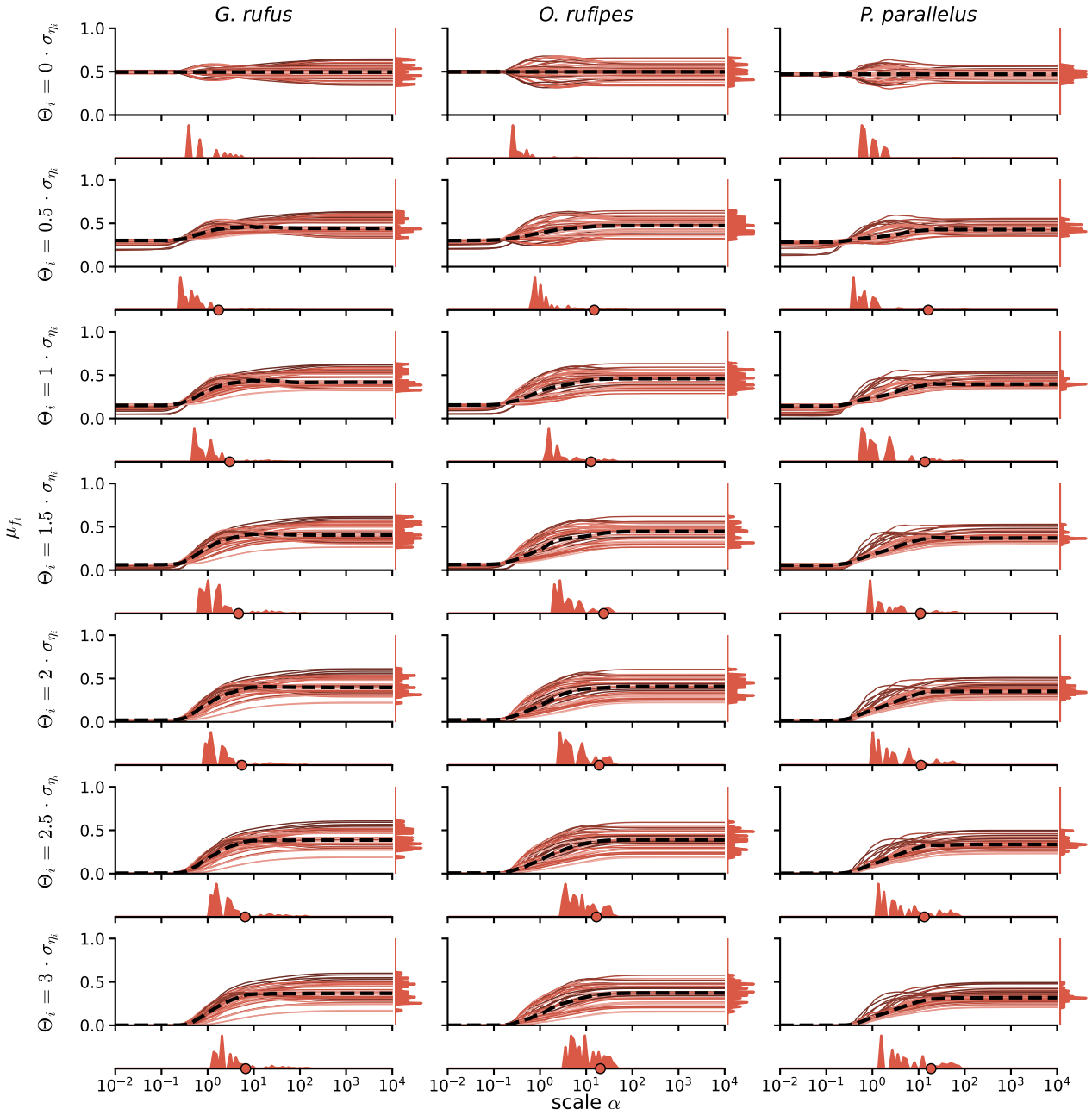
G. rufus
O. rufipes
P. parallelus
$\Theta_i = 0 \cdot \sigma_{\eta_i}$
$\Theta_i = 0.5 \cdot \sigma_{\eta_i}$
$\Theta_i = 1 \cdot \sigma_{\eta_i}$
$\Theta_i = 1.5 \cdot \sigma_{\eta_i}$
$\Theta_i = 2 \cdot \sigma_{\eta_i}$
$\Theta_i = 2.5 \cdot \sigma_{\eta_i}$
$\Theta_i = 3 \cdot \sigma_{\eta_i}$
μ_{f_i}
1.0
0.5
0.0
10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
scale α

Fig. 22: