

Supporting Information

Clemens et al. 10.1073/pnas.1412741111

SI Methods

Stimulus Design. Attractive and nonattractive syllables were created as follows: The attractive syllable consisted of a 72-ms-long noise pulse (5–40 kHz) of constant amplitude followed by 12 ms of silence (Fig. 1B). This type of song subunit exhibits a syllable-to-pause ratio that elicits female responses with high probability (1). The nonattractive syllable was generated by modulating the amplitude of the attractive pulse with Gaussian low-pass noise (0–200 Hz) using a signal-to-noise ratio $\log_{10}(\sigma_{\text{signal}}^2/\sigma_{\text{noise}}^2)$ of ~ 0.5 . This noise modulation of the amplitude introduced gaps in the syllables known to reduce female response probability (2, 3).

Songs consisted of mixed sequences of 33 attractive and nonattractive syllables in which the proportion of attractive syllables varied from 1 to 0 in various parts of the song (Fig. 1C and Fig. S1). The duration of the song models was 2.8 s.

Playback Experiments. Female responses to these songs were quantified in playback experiments. After the first copulation, females of the species *Chorthippus biguttulus* resist multiple matings for a long time, and do not respond to male calling songs. Therefore, we used only virgin females that were placed in a soundproof chamber. Stimulus playback and recording of female responses was controlled by custom-written software (Labview 7; National Instruments) (Matthias Hennig, Humboldt Universität zu Berlin, Berlin, Germany). Each song was presented 18 times in a randomized sequence (for details, see ref. 4).

Female response probability was taken as the fraction of trials that elicited at least one female response song. To account for different motivation levels, probabilities were normalized by the response probability of the stimulus that was most frequently responded to by that female. Females that responded more than three times to 3 s of continuous noise were discarded as non-selective (4/31). For all further analyses, we took the average response probabilities from 27 individual females.

Drift-Diffusion Model. Integration of song was modeled using a drift-diffusion model with sticky bounds:

$$x_{t+1} = \begin{cases} \theta_-, & \text{if } x \leq \theta_- \\ \theta_+, & \text{if } x \geq \theta_+ \\ \frac{x_t}{\tau} + w_+ + \sigma \zeta_t, & \text{if } s_t = 1 \\ \frac{x_t}{\tau} + w_- + \sigma \zeta_t, & \text{if } s_t = -1 \end{cases}$$

The subscript t marks the syllable number—integration is discrete with each syllable corresponding to one time step. x_t is the integrated evidence up to the current syllable (initial value $x_0 = 0$). τ is the integration time constant of the leaky integrator (in units of syllables). s_t is the syllable type and is 1 for attractive and -1 for nonattractive syllables. w_+ and w_- are the weights for each syllable type. ζ_t is zero mean, unit variance Gaussian noise added at each time step, σ sets the SD of that noise. θ_+ and θ_- are the sticky bounds: if the accumulated evidence x_t reaches either threshold, it will stay there and the animal commits to a decision. If no threshold crossing occurs, the animal's decision is based on the sign of the integrated evidence at stimulus end.

We explored alternative formulations of the drift-diffusion model, which lack a noise term or which include a bias term or a time-dependent gain for the sensory evidence (5). Excluding the noise term reduced model performance and led to un-

realistic, binary prediction values. Adding additional parameters (bias term, time-dependent gain) did not improve performance and strongly reduced the reproducibility of the parameter values obtained over individual cross-validations. We therefore used the standard model formulation for all analyses.

In drift-diffusion models, sensory information is provided by dedicated feature detectors for positive and for negative evidence; this is a sensible assumption in situations where an animal is trained to discriminate between two alternatives, e.g., two sound frequencies (6) or the left- or rightward object motion (7). Though there may exist detectors for specific nonattractive features in grasshopper song evaluation (8), explicit recognition of all negative evidence in general is unrealistic. In communication systems in which the majority of signals fail to elicit any response, detecting such a large class of signals would require unfeasibly many feature detectors. Because the integration of evidence in the model is linear, a simple extension can solve this problem: the output of neurons detecting attractive features is combined into a single syllable score. This score is high for attractive syllables and low (i.e., zero) for all nonattractive stimuli. This score can then be mapped onto positive and negative evidence using a negative bias term (Fig. S2).

For this study, we focus on how different types of sensory information—differentiated by the temporal pattern of a song subunit—are integrated to yield the female decision and not on how this integration is initiated. We therefore assume that integration starts at zero and is triggered by the occurrence of a signal that is sufficiently loud against the background noise level and has the adequate, broadband carrier spectrum (1). Focusing on the temporal pattern and not on the carrier spectrum is well justified, because the songs of different grasshopper species in the *Chorthippus* group do not differ much in their carrier spectrum, but in the pattern of amplitude modulations (9, 10).

The model was fitted using a genetic algorithm (11, 12). Because the noisy integration is linear, parameters are determined up to an offset and/or a common scaling factor. We therefore fixed the positive weight to $w_+ = 1$. Our conclusion of asymmetrical integration of evidence is independent of this linear transformation of the parameters. For fitting, the other parameter values were constrained to $[-50, 0]$ for w_- , $[0, 250]$ for τ , $[0, 200]$ for θ_+ , $[-1,000, 0]$ for θ_- . These constraints were found using exploratory trial runs and served to speed up the fitting procedure. We ensured that none of the bounds affected the fitting results.

Model performance was evaluated by leave-one-out cross-validation; i.e., the model was fit using 31 of the 32 stimuli to find the optimal parameter values. Then, a prediction was made for the left-out stimulus. Repeating this procedure such that each stimulus is left out once yields a set of 32 models and one prediction for the left-out stimulus. Model performance was taken as Pearson's coefficient of correlation r^2 between the predicted and measured responses.

Most model parameters were reproducible across the 32 cross-validated models (see SDs in Table 1). Fits starting with different initial parameter sets always converged to similar solutions, indicating that a globally optimal solution was well defined and was always found by the fitting algorithm.

To ensure that model parameters were well constrained by the data, we quantified model performance for parameter values around the optimal one found by the genetic algorithm—if the found optimum lies in an error minimum, then the parameter is well constrained (Fig. S3). All parameters except τ and θ_- lay in

a clearly defined minimum. The integration time constant τ and the negative threshold θ_- were defined only up to a minimal value; as for τ , this was because integration was effectively perfect—hence, as long as τ was much longer than the stimulus duration, model performance was optimal. A model with perfect integration exhibited identical performance and highly similar parameter values (Table 1). As for θ_- , the strongly asymmetrical integration made this parameter only semidetermined—as long as θ_- exceeded a value so that very early noise-induced threshold crossings were improbable, the strong bias induced by the large weight for negative evidence ensured that the model output remained virtually identical. Thus, the fact that two parameters were semidetermined does not reflect a lack of data but the idiosyncrasies of the decision process being modeled.

1. von Helversen O, von Helversen D (1997) Recognition of sex in the acoustic communication of the grasshopper *Chorthippus biguttulus* (Orthoptera, Acrididae). *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 180(4):373–386.
2. Stumpner A, Ronacher B (1994) Neurophysiological aspects of song pattern recognition and sound localization in grasshoppers. *Am Zool* 34(6):696–705.
3. Ronacher B, Stange N (2012) Processing of acoustic signals in grasshoppers: A neuroethological approach towards female choice. *J Physiol Paris* 107(1-2):41–50.
4. Schmidt A, Ronacher B, Hennig RM (2008) The role of frequency, phase and time for processing of amplitude modulated signals by grasshoppers. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 194(3):221–233.
5. Cisek P, Puskas GA, El-Murr S (2009) Decisions in changing conditions: The urgency-gating model. *J Neurosci* 29(37):11560–11571.
6. Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA (2010) Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci* 13(8):1011–1019.
7. Shadlen MN, Kiani R (2013) Decision making as a window on cognition. *Neuron* 80(3):791–806.
8. Ronacher B, Stumpner A (1988) Filtering of behaviourally relevant temporal parameters of a grasshopper's song by an auditory interneuron. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 163(4):517–523.
9. Meyer J, Elsner N (1996) How well are frequency sensitivities of grasshopper ears tuned to species-specific song spectra? *J Exp Biol* 199(Pt 7):1631–1642.
10. Safi, K et al. (2006) Species recognition influences female mate preferences in the common European grasshopper (*Chorthippus biguttulus* Linnaeus, 1758). *Ethology* 112(12):1225–1230.
11. Clemens J, Ronacher B (2013) Feature extraction and integration underlying perceptual decision making during courtship behavior. *J Neurosci* 33(29):12136–12145.
12. Clemens J, Hennig RM (2013) Computational principles underlying the recognition of acoustic signals in insects. *J Comput Neurosci* 35(1):75–85.

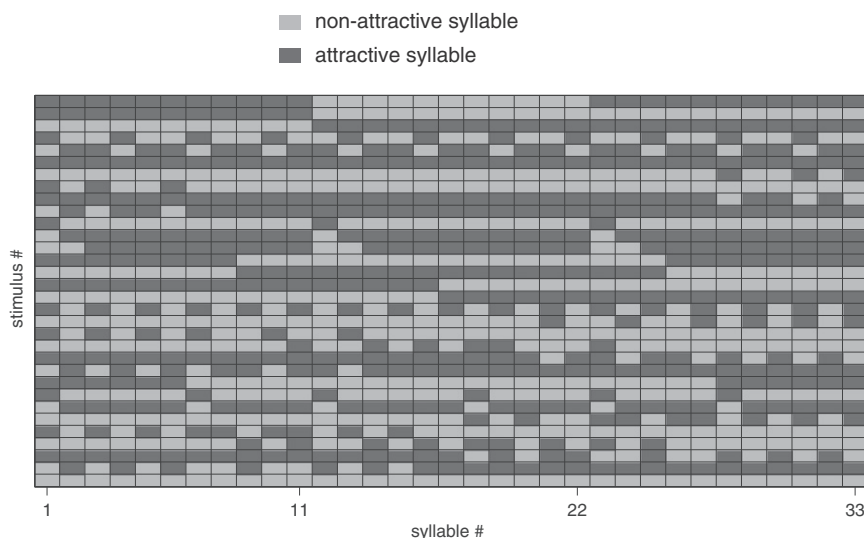


Fig. S1. Stimulus set. Syllable sequences of all 32 songs. Each song consisted of 33 syllables that were either attractive (black) or nonattractive (gray) as shown in (Fig. 1B). Song duration was ~ 2.8 s.

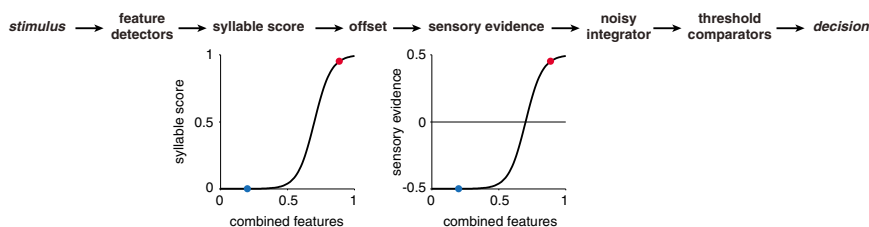


Fig. S2. General model schematic. The pattern of amplitude modulations of a syllable is processed by a bank of feature detectors, the combined output of which yields a score that ranges between 0 (unattractive syllable) and 1 (very attractive syllable). A simple transformation (e.g., a negative offset) turns this score into information with positive and negative values. The information is then accumulated using a noisy integrator and compared with a threshold for a positive or negative response. The blue dot corresponds to a putative nonattractive, the red dot to an attractive syllable.

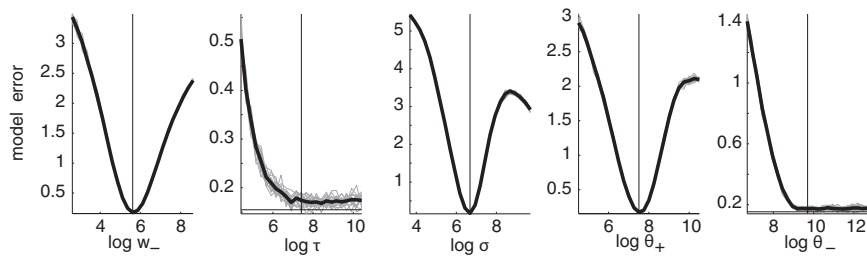


Fig. S3. Model parameters are well constrained by the data. Dependence of model error (normalized mean squared error) on parameter values. The black vertical line indicates the optimal parameter value found by the genetic algorithm. Because the model was stochastic, we estimated the error for 16 runs with independent noise (gray traces) and averaged them (black traces).