Optimal nonlinear cue integration for sound localization

Brian J. Fischer[1] · Jose Luis Peña[2]

Abstract Integration of multiple sensory cues can improve performance in detection and estimation tasks. There is an open theoretical question of the conditions under which linear or nonlinear cue combination is Bayes-optimal. We demonstrate that a neural population decoded by a population vector requires nonlinear cue combination to approximate Bayesian inference. Specifically, if cues are conditionally independent, multiplicative cue combination is optimal for the population vector. The model was tested on neural and behavioral responses in the barn owl's sound localization system where space-specific neurons owe their selectivity to multiplicative tuning to sound localization cues interaural phase (IPD) and level (ILD) differences. We found that IPD and ILD cues are approximately conditionally independent. As a result, the multiplicative combination selectivity to IPD and ILD of midbrain space-specific neurons permits a population vector to perform Bayesian cue combination. We further show that this model describes the owl's localization behavior in azimuth and elevation. This work provides theoretical justification and experimental evidence supporting the optimality of nonlinear cue combination.

Action Editor: Susanne Schreiber

✉ Brian J. Fischer
fischer9@seattleu.edu

[1] Department of Mathematics, Seattle University, 901 12th Ave, Seattle, WA 98122, USA

[2] Department of Neuroscience, Albert Einstein College of Medicine, 1410 Pelham Parkway South, Bronx, NY 10461, USA

1 Introduction

Perception in natural environments depends on the integration of multiple sensory cues within and across modalities. When sensory cues provide complementary information or are corrupted by independent noise, combining them can lead to improved performance (van Beers et al. 1999; Ernst and Banks 2002; Battaglia et al. 2003; Oruç et al. 2003; Alais and Burr 2004; Hillis et al. 2004; Knill and Pouget 2004; Gu et al. 2008; Landy et al. 2011; Ohshiro et al. 2011; Hollensteiner et al. 2015). This occurs, for example, when visual and auditory cues are combined to determine the position of an object (Battaglia et al. 2003; Alais and Burr 2004; Whitchurch and Takahashi 2006). Performance improvement through cue combination can also occur within a single modality. For example, multiple visual cues are combined for depth perception (Landy et al. 1995; Jacobs 1999) and multiple auditory cues are combined for sound localization (Moiseff 1989; Middlebrooks and Green 1991; Wightman and Kistler 1993) or auditory scene analysis (Bregman 1994; Roman et al. 2003). Although the performance benefit of cue combination is well understood, whether cue combination by neural populations approaches optimality is an open question.

Bayesian inference provides a framework for studying how multiple sensory cues and prior information about the environment can be integrated optimally to drive behavior (Knill and Pouget 2004; Angelaki et al. 2009). Bayesian inference relies on a posterior distribution, which describes what is known about the environment given the sensory input. In many studies of sensory-cue combination, cues are assumed to be conditionally independent (e.g., van Beers et al. 1999; Ernst and Banks 2002; Battaglia et al. 2003; Jacobs 1999; Alais and Burr 2004; Hillis et al. 2004). Conditional independence of sensory cues means that the probability of one cue does not depend on the value of the other, given the state of the environment. Bayes' theorem

specifies that the posterior distribution, given conditionally independent sensory inputs, is determined by a product of sensory-cue likelihoods and the prior over the environmental variables.

Previous studies have used linear and nonlinear approaches to model Bayes-optimal cue combination in neural circuits. Several proposals suggest that the neural basis of optimal cue combination for Bayesian inference is a linear combination of neural responses (Ma et al. 2006; Fetsch et al. 2011). Multiplying the likelihood functions in Bayes' theorem can correspond to a linear combination of firing rates when the distributions of neural responses are Poisson-like (Ma et al. 2006; Beck et al. 2007). Alternatively, a linear combination of firing rates may also be optimal if firing rates match the log-likelihood function, where the product of the likelihoods is achieved through the sum of log-likelihoods. Modeling work has shown that a network may compute such an operation with a linear combination of firing rates (Jazayeri and Movshon 2006). Nonlinear Bayes-optimal cue combination models can be divided into classes of models. For one, models that utilize a hidden layer of neurons acting as basis functions for nonlinear computation (Eliasmith and Anderson 2004; Beck et al. 2011). These models rely on the presence of neurons with nonlinear tuning to the cues that allow the network to perform accurate function approximation (Eliasmith and Anderson 2004) or information preservation (Beck et al. 2011) but the specific form of nonlinearity remains undefined. A second alternative class of nonlinear models for the neural implementation of Bayesian inference takes into account the ubiquitous non-uniformity in neural representations. In these models, referred to as non-uniform population codes, the statistical structure of the environment is encoded in the non-uniform distribution of preferred stimuli and tuning curve shapes across the population. Studies in birds and humans are consistent with this theory (Fischer and Peña 2011; Girshick et al. 2011; Cazettes et al. 2014). These studies have shown that the statistical relationship between environmental variables (such as image boundaries or the location of a sound source) and the sensory cues used to make inferences about these variables (edge orientation (Girshick et al. 2011) and interaural time difference (ITD) (Shi and Griffiths 2009; Fischer and Peña 2011), respectively) could be represented in the non-uniform tuning properties. In natural environments, the visual and auditory tasks of scene segmentation and sound localization may rely on the integration of multiple sensory cues (Moiseff 1989; Landy and Kojima 2001). However, the question of what operations neurons perform for Bayes-optimal cue combination in the non-uniform population code model also remains open.

Here we investigate the neural basis of optimal cue combination in the owl's sound localization system. To localize sounds, animals rely on the noisy and ambiguous sensory cues ITD and interaural level difference (ILD) (Knudsen and Konishi 1979; Moiseff 1989; Brainard et al. 1992). In the owl's external nucleus of the inferior colliculus (ICx), and its direct projection site the optic tectum (OT), there is a map of auditory

space (Knudsen and Konishi 1978; Knudsen 1982). It has been shown that the spatial selectivity of ICx and OT neurons depends on the tuning to ITD and ILD, where ITD is used for azimuth and ILD for elevation (Moiseff and Konishi 1983; Moiseff 1989; Brainard et al. 1992). The combination selectivity to ITD and ILD has been shown to emerge by an effective multiplication of inputs tuned to ITD and ILD (Peña and Konishi 2001; Fischer et al. 2007). Estimating sound direction from ITD and ILD involves several sources of uncertainty. First, the ITD and ILD computed in the brain are corrupted by noise from the environment (Spitzer et al. 2003; Cazettes et al. 2014) and neural computation (Christianson and Peña 2006; Pecka et al. 2010). Second, ITD and ILD are not uniquely related to a particular direction in space (Brainard et al. 1992; Fischer and Peña 2011). Additional ambiguity arises because the true ITD cannot be distinguished from other ITDs corresponding to equivalent interaural phase differences (IPD) in the owl's early sound localization pathway, where ITD is computed in narrow frequency bands (Wagner et al. 1987; Brainard et al. 1992; Peña and Konishi 2000).

The owl's reliance on sound localization for survival (Konishi 1993) and the observation that the owl's localization behavior in the horizontal dimension is consistent with Bayesian inference (Fischer and Peña 2011) motivates the hypothesis that neurons in the owl's localization pathway should combine the spatial cues IPD and ILD optimally. Furthermore, the multiplicative integration of ITD and ILD tunings underlying the spatial selectivity of the owl's space-specific neurons (Peña and Konishi 2001) offers the opportunity to examine the role of nonlinear integration in optimal cue combination. Here we derive the optimal form of cue combination in a non-uniform population code that matches the statistics of the environment and is read-out by a population vector (PV). We then show that IPD and ILD cues are approximately conditionally independent and that the owl's localization behavior in the horizontal and vertical dimensions is described by an optimal combination of these cues (Fig. 1).

2 Materials and methods

2.1 IPD and ILD estimated from head-related transfer functions (HRTFs)

The HRTFs of ten barn owls were provided by Dr. Keller (Keller et al. 1998) from the University of Oregon. The ILD and IPD cues were computed from the HRTFs by first convolving a white noise stimulus (0.5–12 kHz) with the head-related impulse responses for the left and right ears at the target direction. The left and right ear inputs were then filtered with a gammatone filterbank having center frequencies covering 2–9 kHz, equal gains across frequency, and bandwidths estimated from barn owl
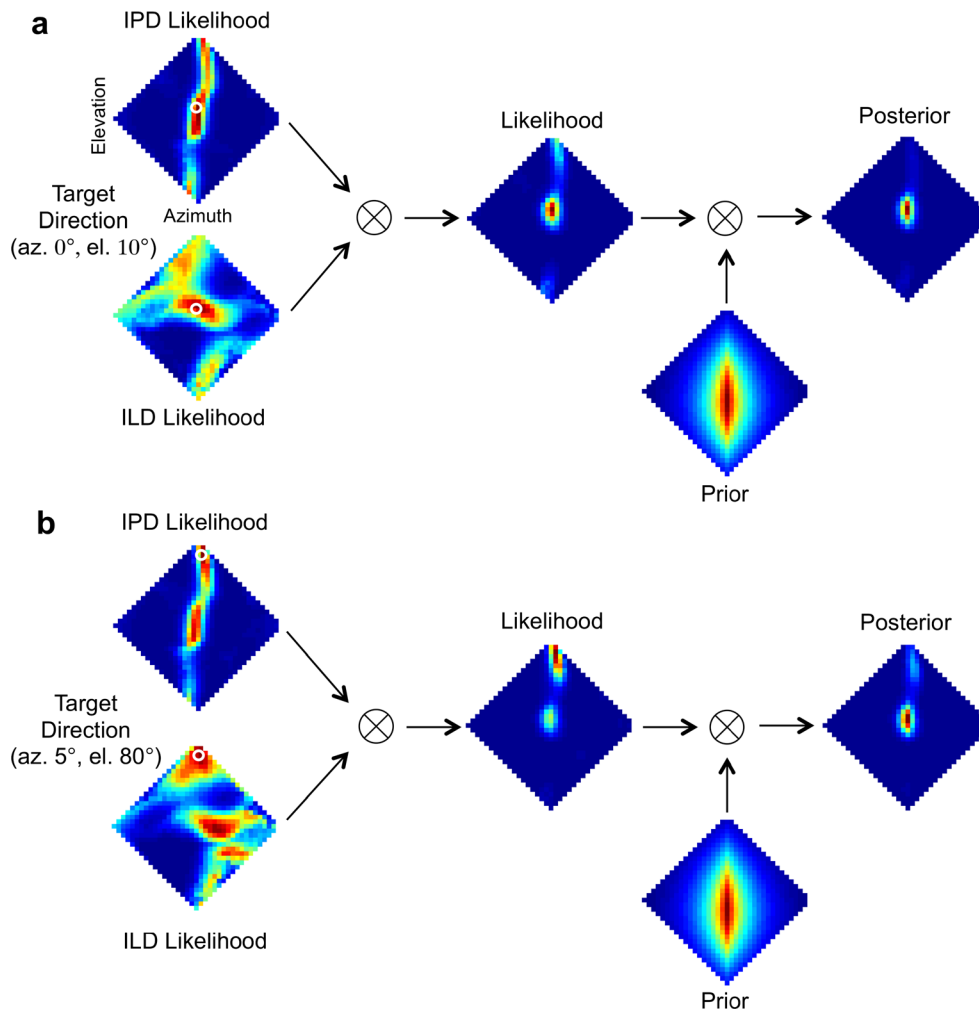
Fig. 1 Bayesian cue combination. **a, b** Components of the Bayesian model for two example stimulus directions **a** (0°,10°) and **b** (5°,80°). The *diamond plots* show the frontal hemisphere measured in double-polar coordinates corresponding to azimuth (*horizontal*) and elevation (*vertical*) directions. Azimuth and elevation each change in five-degree steps. The color scale ranges from zero (*blue*) to the maximum of the plotted function (*red*). ILD and IPD provide complementary information about sound location. Left, the IPD likelihood is primarily restricted in azimuth and the ILD likelihood is primarily restricted in elevation. The target direction is indicated by a *white circle*. Center top, the ILD-IPD likelihood is a product of each cue's likelihood. It has a peak at the true source direction, but also has secondary peaks. Center bottom, the prior emphasizes directions at the center in azimuth and below the center in elevation. Right, the posterior is the product of the likelihood and prior and has a single dominant peak. For a source direction near the center (**a**), the posterior is more focused on the true source direction than is the likelihood. In contrast, for a source direction in the periphery (**b**), the posterior is biased away from the source direction toward the center of gaze

auditory-nerve fiber responses (Köppl 1997b). IPD was calculated in each frequency channel from the delay at the peak of the cross-correlation of the left and right filterbank outputs. ILD was calculated in each frequency channel as the difference in the logarithms of the root-mean-square values of the filterbank outputs.

2.2 Bayesian model of sound localization

To define the likelihood, we consider a model where the sensory observation made by the owl is given by the ILD and IPD spectra derived from barn owl head-related transfer functions (HRTFs) after the spectra are corrupted with additive noise. For a source azimuth θ and elevation ϕ, the observation vector $\mathbf{s}$ is expressed as

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_{\text{IPD}} \\ \mathbf{s}_{\text{ILD}} \end{bmatrix} = \begin{bmatrix} \mathbf{IPD}_{\theta,\phi} \\ \mathbf{ILD}_{\theta,\phi} \end{bmatrix} + \begin{bmatrix} \eta_{\text{IPD}} \\ \eta_{\text{ILD}} \end{bmatrix} \qquad (1)$$

where the ILD spectrum $\mathbf{ILD}_{\theta,\phi} = [\text{ILD}_{\theta,\phi}(\omega_1), \text{ILD}_{\theta,\phi}(\omega_2), \ldots, \text{ILD}_{\theta,\phi}(\omega_K)]$ and the IPD spectrum $\mathbf{IPD}_{\theta,\phi} = [\text{IPD}_{\theta,\phi}(\omega_1), \text{IPD}_{\theta,\phi}(\omega_2), \ldots, \text{IPD}_{\theta,\phi}(\omega_K)]$ are specified at frequencies ω_i between 3 and 9 kHz in steps of 0.6 kHz.

The noise corrupting the ILD spectrum η_{ILD} is modeled as a Gaussian random vector with independent components. The variance of each component is frequency- and direction-dependent. The IPD noise η_{IPD} is assumed to have a circular

Gaussian distribution with mean zero at each frequency. As is the case for ILD, the variance of each IPD noise component is frequency- and direction-dependent. We assume that the ILD and IPD noise terms η_{ILD} and η_{IPD} are mutually independent conditioned on the source direction.

We calculated the environmental variability of IPD and ILD over different directions of concurrent noise sources (Cazettes et al. 2014). For each target direction, a white noise stimulus (0.5–12 kHz) was convolved with the head-related impulse responses for the left and right ears at the target direction. The directions of the target sounds covered the frontal hemisphere using five-degree steps measured in double polar coordinates, leading to 685 target directions. Additionally, concurrent white noise sources were simulated as arising from directions surrounding the owl by convolving the noise with head-related impulse responses for the left and right ears at the direction of the second source. The directions of concurrent noise sources covered directions surrounding the owl using all possible elevations in the frontal hemisphere and elevations between −25 and 25° for directions in the rear hemisphere, leading to 952 total directions. The signals from the two sources were added together to form the input to the left and right ears. The left and right ear inputs were then filtered with a gammatone filterbank having center frequencies covering 2–9 kHz, equal gains across frequency, and bandwidths estimated from barn owl auditory-nerve fiber responses, as above (Köppl 1997b). We calculated the IPD $\sigma^2_{IPD,\theta,\phi}(\omega_i)$ and ILD variances $\sigma^2_{ILD,\theta,\phi}(\omega_i)$ and the correlation between IPD and ILD over all directions of the second noise source. These variances are a function of the azimuth θ, elevation ϕ, and frequency ω_i of the target sound.

The likelihood function of azimuth θ and elevation ϕ for observed cues $\mathbf{s}_{IPD}$ and $\mathbf{s}_{ILD}$ has the form

$$p_{\mathbf{s}|\Theta,\Phi}(\mathbf{s}|\theta,\phi) = p_{\mathbf{s}_{IPD}|\Theta,\Phi}(\mathbf{s}_{IPD}|\theta,\phi)p_{\mathbf{s}_{ILD}|\Theta,\Phi}(\mathbf{s}_{ILD}|\theta,\phi) \qquad (2)$$

where the ILD likelihood function is a Gaussian given by

$$p_{\mathbf{s}_{ILD}|\Theta,\Phi}(\mathbf{s}_{ILD}|\theta,\phi) \propto \exp\left[-\frac{1}{2}\sum_{j=1}^{N}\left(\frac{s_{ILD}(\omega_j)-ILD_{\theta,\phi}(\omega_j)}{\sigma_{\theta,\phi}(\omega_j)}\right)^2\right] \quad (3)$$

and the IPD likelihood function is a circular Gaussian given by

$$p_{\mathbf{s}_{IPD}|\Theta,\Phi}(\mathbf{s}_{IPD}|\theta,\phi) \propto \exp\left[\sum_{j=1}^{N}\kappa_{\theta,\phi}(\omega_j)\cos(s_{IPD}(\omega_j)-IPD_{\theta,\phi}(\omega_j))\right] \quad (4)$$

where $\kappa_{\theta,\phi}(\omega_j)$ is a direction- and frequency-dependent parameter that determines the variance. We assume that the overall variance in IPD and ILD is a sum of the variance due to concurrent sources and a constant variance due to noise in neural computations: $\sigma^2_{\theta,\phi}(\omega_i) = \sigma^2_{ILD,\theta,\phi}(\omega_i) + v_{ILD}$ and $\frac{1}{\kappa_{\theta,\phi}(\omega_j)} = \sigma^2_{IPD,\theta,\phi}(\omega_i) + v_{IPD}$. The constant variances due to noise in neural computations v_{ILD} and v_{IPD} do not depend on direction or frequency and are parameters in

the model that are found by fitting the model to the owl's localization behavior. Behavioral data were taken from published reports in (Knudsen et al. 1979) of the absolute angular error for two owls performing head turns to sounds presented from speakers.

The prior density $p_{\Theta,\Phi}(\theta,\phi)$ is proportional to the product of an elevation component $p_\Phi(\phi)$ and an azimuth component $p_\Theta(\theta)$: $p_{\Theta,\Phi}(\theta,\phi) \propto p_\Theta(\theta)p_\Phi(\phi)$. The elevation component of the prior density is a combination of two Gaussian functions that may have different widths above and below the mode μ_ϕ:

$$p_\Phi(\phi) = \begin{cases} \exp\left(-\frac{1}{2\sigma^2_{\phi_1}}\left(\phi-\mu_\phi\right)^2\right) & \phi \le \mu_\phi \\ \exp\left(-\frac{1}{2\sigma^2_{\phi_2}}\left(\phi-\mu_\phi\right)^2\right) & \phi > \mu_\phi \end{cases}. \qquad (5)$$

The azimuth component of the prior density is a Laplace density

$$p_\Theta(\theta) \propto \exp\left(-\frac{|\theta|}{\beta}\right), \qquad (6)$$

with variance $2\beta^2$. Parameters of the prior density were found by fitting the Bayesian model to the owl's localization behavior.

The Bayesian estimate of stimulus direction in azimuth θ and elevation ϕ from the noisy IPD and ILD spectra is given by the mean direction under the posterior distribution $p_{\Theta,\Phi|\mathbf{s}}(\theta,\phi|\mathbf{s})$. The mean direction is found by first computing the vector BV that points in the mean direction as

$$BV = \iint u(\theta,\phi)p_{\Theta,\Phi|\mathbf{s}}(\theta,\phi|\mathbf{s})d\theta d\phi \propto \iint u(\theta,\phi)p_{\mathbf{s}|\Theta,\Phi}(\mathbf{s}|\theta,\phi)p_{\Theta,\Phi|\mathbf{s}}(\theta,\phi|\mathbf{s})d\theta d\phi$$
$$(7)$$

where $u(\theta,\phi)$ is a unit vector pointing in direction (θ,ϕ) and the proportionality follows from Bayes' rule. The direction estimate is the direction of the mean vector.

3 Conditional independence of IPD and ILD

We analyzed the conditional independence of IPD and ILD by comparing kernel density estimates of the full joint distribution and the joint distribution assuming conditional independence. The kernel density estimate of the full joint distribution at one frequency from observations $\{IPD_n(\omega_j), ILD_n(\omega_j)\}_{n=1}^N$ was given by

$$k(IPD, ILD|\theta,\phi) \propto \sum_{n=1}^{N}\exp\left[\alpha\cos(IPD-IPD_n(\omega_j))\right] \qquad (8)$$
$$\exp\left[-\frac{1}{2\beta^2}\left(ILD-ILD_n(\omega_j)\right)^2\right]$$

and the kernel density estimate assuming conditional independence was given by

$$k_{ind}(\text{IPD}, \text{ILD}|\theta, \phi) \propto \sum_{n=1}^{N} \exp\left[\alpha\cos\left(\text{IPD}-\text{IPD}_n(\omega_j)\right)\right] \sum_{n=1}^{N} \quad (9)$$

$$\exp\left[-\frac{1}{2\beta^2}\left(\text{ILD}-\text{ILD}_n(\omega_j)\right)^2\right]$$

The parameters $\alpha = 11$ and $\beta = 1$ were selected so that the IPD and ILD components had widths that were the same percentage of the range of IPD and ILD, respectively. The similarity of the kernel density estimates was assessed using the Kullback–Leibler divergence (Kullback and Leibler 1951) and the fractional energy in the first singular value of the singular value decomposition of the joint density.

4 Neural model

The neural model consisted of a population of 1000 direction-selective neurons that model the OT. The model population size is small relative to the number of neurons in OT (Knudsen 1983). The preferred directions were drawn independently from the prior distribution over direction $p_{\Theta,\Phi}(\theta, \phi)$.

The neural tuning curves are proportional to the likelihood function and are given by

$$f_n(\mathbf{s}_{\text{IPD}}, \mathbf{s}_{\text{ILD}}) = f_{\max} p_{\mathbf{s}_{\text{IPD}}|\Theta,\Phi}(\mathbf{s}_{\text{IPD}}|\theta_n, \phi_n) p_{\mathbf{s}_{\text{ILD}}|\Theta,\Phi}(\mathbf{s}_{\text{ILD}}|\theta_n, \phi_n). \quad (10)$$

The maximum firing rate was set to 10 spikes/stimulus (Saberi et al. 1998). During simulations, the neurons have independent Poisson distributed firing rates $r_n(\mathbf{s}_{\text{IPD}}, \mathbf{s}_{\text{ILD}})$ with mean values given by the neural tuning curves $f_n(\mathbf{s}_{\text{IPD}}, \mathbf{s}_{\text{ILD}})$.

5 Population vector

The population vector is computed as a linear combination of the preferred direction vectors of the neurons, weighted by the firing rates

$$PV(\mathbf{s}_{\text{IPD}}, \mathbf{s}_{\text{ILD}}) = \frac{1}{N}\sum_{n=1}^{N} u(\theta_n, \phi_n) r_n(\mathbf{s}_{\text{IPD}}, \mathbf{s}_{\text{ILD}}) \quad (11)$$

where $u(\theta_n, \phi_n)$ is a unit vector pointing in the nth neuron's preferred direction. The direction estimate is the direction of the population vector (PV).

6 Density of preferred directions in the midbrain

To determine the density of preferred directions in OT, we used measurements of the shape of the OT auditory space map (Knudsen 1982), as described previously for azimuth (Fischer and Peña 2011). Briefly, assuming that cell density

is homogeneous in OT, the physical distance between points corresponding to different preferred directions in the auditory space map will be proportional to the number of cells that lie between those directions. The relationship between preferred elevation and position in the auditory-space map may be described by a curve that is proportional to the cumulative distribution function of the density of preferred directions. To estimate the density of preferred elevations, we fit the relationship between preferred elevation and position in the OT space map with a curve that is proportional to the cumulative distribution function of a piecewise Gaussian density as used in the Bayesian model (Eq. 5). We fit the relationship between preferred azimuth and position in the OT space map with a curve that is proportional to a cumulative Laplace distribution function. The overall density of preferred directions in azimuth is a mixture of the fitted Laplace density and its mirror image for the other hemisphere (Fischer and Peña 2011).

7 Extracellular recording

Methods for surgery, stimulus delivery, and data collection have been described previously (Fischer et al. 2007). Briefly, four barn owls (*Tyto alba*) were anesthetized with intramuscular injections of ketamine (20 mg/kg; Ketaject; Phoenix Pharmaceuticals, St. Joseph, MO) and xylazine (2 mg/kg; Xyla-Ject; Phoenix Pharmaceuticals). Extracellular recordings of single ICcl neurons ($n = 77$) were made with tungsten electrodes (1 MΩ, 0.005-in.; A-M Systems, Carlsborg, WA). All recordings took place in a double-walled sound-attenuating chamber (Industrial Acoustics, Bronx, NY). Acoustic stimuli were delivered by a stereo analog interface [DD1; Tucker Davis Technologies (TDT), Gainesville, FL] through a calibrated earphone assembly. Stimuli for both intracellular and extracellular recordings consisted of broadband noise (0.5–12 kHz) 100 ms in duration with 5-ms linear rise and fall ramps. Stimulus ILD was varied in steps of 3–5 dB.

8 Intracellular recording

Methods for *in vivo* intracellular recordings of ICx neurons ($n = 12$) were described in detail and published previously (Peña and Konishi 2001, 2002, 2004). Briefly, barn owls were anesthetized by intramuscular injection of ketamine hydrochloride (25 mg/kg; Ketaset; Phoenix Pharmaceuticals, Mountain View, CA) and diazepam (1.3 mg/kg; Steris Laboratories, Phoenix, AZ). ICx was approached through a hole made on the exoccipital bone, which provided easier access to the optic lobe. All experiments were performed in a double-walled sound-attenuating chamber.

Sharp borosilicate glass electrodes filled with 2 M potassium acetate and 4 % neurobiotin were used. Analog signals

were amplified (Axoclamp 2A) and stored in the computer. The tracer neurobiotin was injected by iontophoresis at the end of the recording (3 nA positive 300 ms current steps, 3 per second for 5 to 30 min). After the experiment owls were overdosed with Nembutal and perfused with 2 % paraformaldehyde. Brain tissue was cut in 60 μm thick sections and processed according to standard protocols (Kita and Armstrong, 1991).

We computed the median membrane potential during the first 50 ms of the response to sound and averaged it over three to five stimulus presentations. Mean resting potentials are the means of median membrane potentials averaged over all trials within a period of 100 ms before each stimulus onset. ITD and intensity response curves of median membrane potential responses were made by custom software written in Matlab.

9 Sound stimulation

Acoustic stimuli were digitally synthesized by a computer and delivered to both ears through calibrated earphones. Auditory stimuli consisted of broadband noise bursts (0.5–12.0 kHz; 50–100 ms duration and 5 ms rise and decay times; sound level was 40–50 dB sound pressure level. The computer synthesized three random signals to obtain different values of binaural correlation. One noise signal was delivered to both ears, making for the correlated component of the sound. The other two signals were used as the uncorrelated component of the stimulus by adding them to the correlated sound in varying amounts, while keeping the sound level constant. Binaural correlation varies with the relative amplitude of the uncorrelated and correlated noises by $1/(1 + k^2)$, where k is the ratio between the root-mean-square amplitudes of the uncorrelated and correlated noises.

10 Results

Below, we first use a theoretical approach to demonstrate that nonlinear combination of sensory cues allows for optimal Bayesian estimation provided cues are conditionally independent. Next, we confirm that neural responses perform nonlinear cue combination and that ITD and ILD are approximately conditionally independent. Finally, we show that the owl's localization behavior in azimuth and elevation is described by the Bayesian model.

11 Nonlinear integration supports optimal cue integration

We first examine whether in this particular Bayesian framework optimal cue integration within (unisensory)

or across (multisensory) sensory modalities is nonlinear. Consider the problem of inferring the value of an environmental stimulus X from two sets of cues C_1 and C_2. The cues are ultimately encoded by neurons that have preferred stimuli X_n and mean responses to the cues given by $f_n(C_1, C_2)$. An example multisensory problem of this type is to infer the position of an object from auditory and visual cues. An example unisensory problem is to infer the position of a sound source from IPD and ILD. The Bayesian solution to this inference problem is to estimate the value of X from the posterior probability $p_{X|C_1,C_2}$ $(X|C_1, C_2)$. If the preferred directions of neurons in the population X_n are drawn from the prior distribution $p_X(X)$ and the population response is proportional to the likelihood $f_n(C_1, C_2) \propto p_{C_1,C_2|X}(C_1, C_2|X_n)$, then a readout across the population with a PV will accurately approximate the mean of the posterior (Shi and Griffiths 2009; Fischer and Peña 2011). This analysis predicts that optimal cue combination is nonlinear because the likelihood will be a nonlinear function of the cues: $f_n(C_1, C_2) \propto p_{C_1|X}$ $(C_1|X_n)p_{C_2|X,C_1}(C_2|X_n, C_1) = g(C_1)h(C_1, C_2)$. In general, cue combination can improve performance the most when cues have independent noise or provide different pieces of information about the stimulus. If the cues C_1 and C_2 are conditionally independent given X, then the likelihood factors and optimal cue combination is multiplicative:

$$f_n(C_1, C_2) \propto p_{C_1|X}(C_1|X_n)p_{C_2|X}(C_2|X_n) = g_n(C_1)h_n(C_2).$$

Thus the main theoretical result is that for the PV to accurately approximate a Bayesian estimate, cue combination must be nonlinear. In particular, multiplicative neural responses are optimal for the combination of conditionally independent cues.

We now interpret this result in the context of estimating the azimuth and elevation (θ, ϕ) from IPD and ILD, and specify the combination rule that is optimal for neurons commanding this task, in the owl's ICx. Because the spatial selectivity of ICx neurons results from their tuning to IPD and ILD (Moiseff and Konishi 1983; Brainard et al. 1992), the response of an ICx neuron with preferred direction (θ_n, ϕ_n) to IPD and ILD can be described by a tuning function $f_n(\text{IPD}, \text{ILD})$, where tuning to IPD and ILD is different for neurons with different preferred directions. If the sensory cues IPD and ILD are conditionally independent given the stimulus direction, then the likelihood function factors as a product of an IPD-based likelihood and an ILD-based likelihood: $p_{s|\Theta,\Phi}(\text{IPD}, \text{ILD}|\theta, \phi) = p_{s_{\text{IPD}}|\Theta,\Phi}(\text{IPD}|\theta, \phi)p_{s_{\text{ILD}}|\Theta,\Phi}(\text{ILD}|\theta, \phi)$. In the non-uniform population code model we consider for the neural implementation of Bayesian inference (Shi and Griffiths 2009; Fischer and Peña 2011), the optimal neural representation of the sensory statistics requires

tuning curves for IPD and ILD that are proportional to the likelihood function: $f_n(\text{IPD}, \text{ILD}) \propto p_{\mathbf{s}|\Theta,\Phi}(\text{IPD}, \text{ILD}|\theta_n, \phi_n)$. Thus, optimal cue combination is given by a product of one function of IPD and one function of ILD:

$$f_n(\text{IPD}, \text{ILD}) \propto p_{\mathbf{s}_{\text{IPD}}|\Theta,\Phi}(\text{IPD}|\theta_n, \phi_n) p_{\mathbf{s}_{\text{ILD}}|\Theta,\Phi}(\text{ILD}|\theta_n, \phi_n) = g_n(\text{IPD})h_n(\text{ILD}).$$

## 12 Statistics of spatial cues used for sound localization

Testing the Bayesian model for optimal combination of IPD and ILD cues requires a description of the variability in the sensory input and neural computations determining the likelihood functions of IPD and ILD. For this, we considered the sensory cues used by the owl for sound localization, given by the IPD and ILD spectra derived from barn owl head-related transfer functions (HRTFs), and how these cues can be corrupted by noise. The relationship between directions in space and IPD and ILD is known to be highly frequency dependent (Brainard and Knudsen 1993; Keller et al. 1998). The relationship is also ambiguous, where IPD and ILD cues near the center of gaze may be similar to those above and below the owl on the vertical plane (Fig. 2; (Brainard et al. 1992)). In addition, the sound localization cues that the owl uses to infer the source direction are subject to variability due to the nature of the sound, the presence of background noise (Nix and Hohmann 2006; Cazettes et al. 2014), and noise in neural computation (Christianson and Peña 2006; Fischer and Konishi 2008). Neural noise is particularly important for IPD because it limits the frequency range over which IPD cues are useful for sound localization. While IPD is a well-defined parameter of the acoustic signals over the entire audible frequency range in any animal, it is only useful in practice for sound localization up to approximately 9 kHz in barn owls and 5 kHz or lower in mammals because of limits in the abilities of neurons to phase lock to the stimulus at high frequencies (Johnson 1980; Palmer and Russell 1986; Köppl 1997a).

We used owl HRTFs ($n = 10$, kindly provided by Clifford Keller) to determine the form of the variability in IPD and ILD as a function of frequency for different sound source directions, assuming naturalistic environments where concurrent sounds are usually present. We computed the variability of IPD for directions covering the frontal hemisphere, as previously shown on the horizontal plane (Cazettes et al. 2014), and expanded the analysis to include variability of ILD cues over different directions of concurrent noise sources. We calculated the reliability of ITD and ILD, defined as the inverse of the variance of each cue. The reliability of both IPD and ILD was largest for central directions at frequencies above approximately 4 kHz (Fig. 3). At lower frequencies, the reliability of ITD and ILD was less spatially dependent.
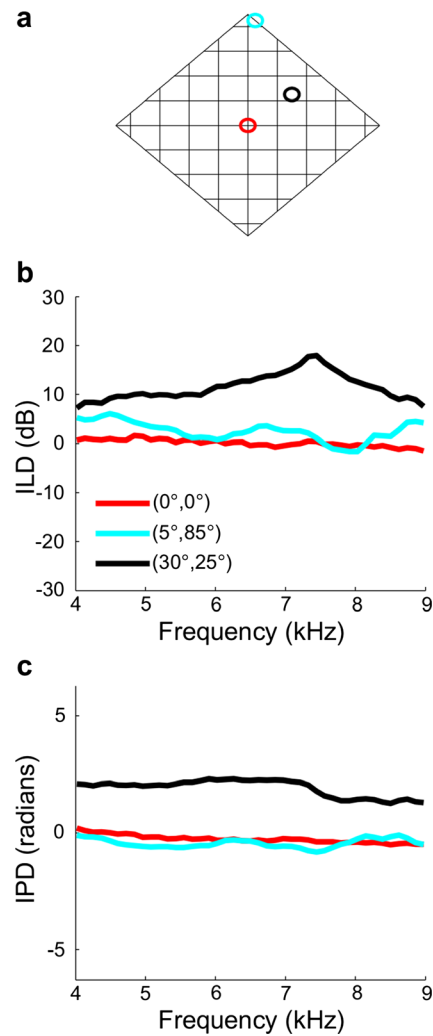
Fig. 2 Ambiguity of sound localization cues. **a** Sounds from three different directions on the vertical plane (*cyan*, (5°,85°), *red*, (0°,0°) and *black* (30°,25°)). **b** ILD spectra for sound directions in (**a**). **c** IPD spectra for directions in (**a**). While the peripheral direction (*cyan*) has similar ILD and IPD spectra than sounds from a frontal direction (*black*), a source at an intermediate location (*black*) can have very different ILD and IPD spectra

IPD and ILD are also subject to variability due to the neural computation underlying the emergence of the tuning for these cues. We used extracellular recordings of neural responses to IPD and ILD to assess this variability. Because the selectivity to both IPD and ILD is created by processing in narrow frequency channels (Manley et al. 1988; Carr and Konishi 1990; Mogdans and Knudsen 1994), we analyzed the tuning variability to IPD and ILD on a frequency-by-frequency basis. A previous study found that the variability of IPD tuning was constant across frequency channels in the owl's localization pathway (Cazettes et al. 2014), indicated by a lack of correlation between the Fano factor of responses to IPD and stimulus frequency in ICx. Here, we extended this analysis to ILD coding by examining the neural
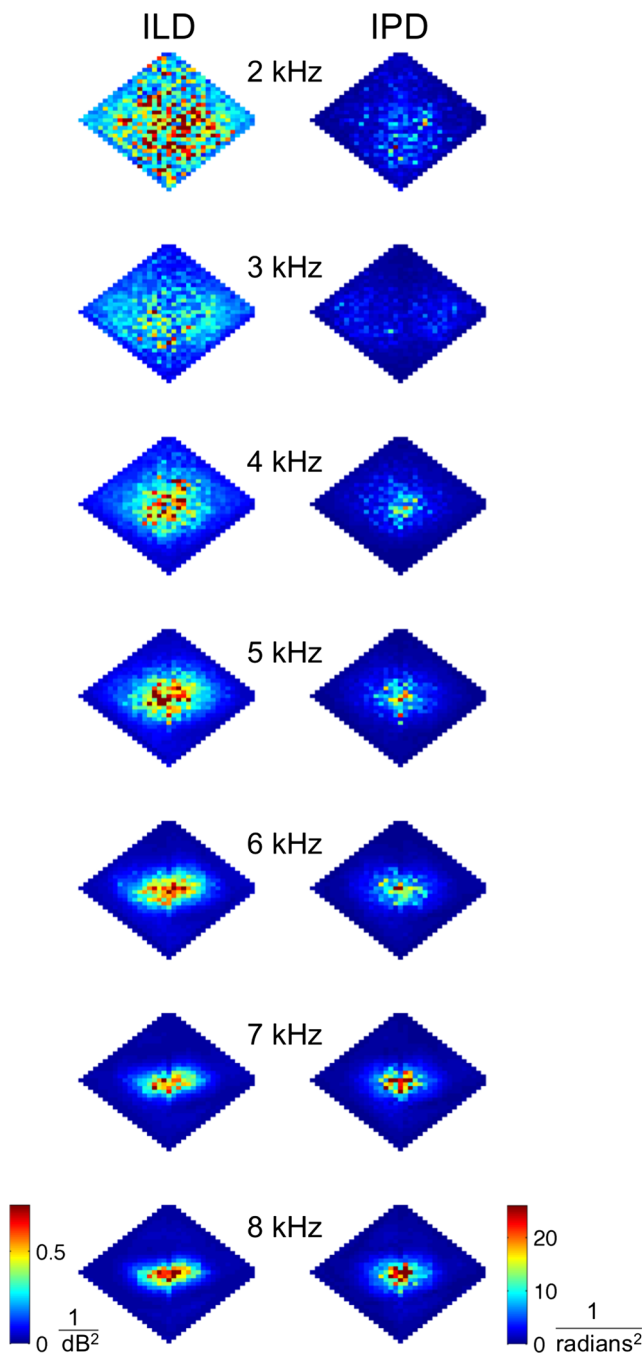
Fig. 3 Reliability of sound localization cues ILD and IPD. The reliability (inverse variance) of ILD (*left*) and IPD (*right*) at each target direction in the frontal hemisphere in the presence of concurrent sources from other directions is plotted separately for frequencies between 2 and 8 kHz. The color axis is on the same scale for all ILD and IPD plots. Overall, the reliability of both IPD and ILD was highest for central directions at high frequencies

variability of ILD tuning in the lateral shell of the central nucleus of the inferior colliculus (ICcl), which projects directly to ICx (Knudsen 1983). ICcl neurons are tuned to ITD and ILD but, unlike ICx, they are narrowly tuned to frequency. We used the average Fano factor over ILD

to quantify the variability of ILD tuning of each ICcl neuron (min = 0.13, median = 0.77, max = 2.55, $n = 77$). There was no significant correlation between the best frequency and the average Fano factor in the sample of ICcl neurons with best frequencies ranging from 500 to 7900 Hz ($r = 0.11$, $p = 0.39$, $n = 77$). Therefore, we assumed a frequency-independent level of variability in IPD and ILD due to neural computation.

13 Conditional independence of IPD and ILD

Our theoretical result specifies that multiplicative cue combination is optimal for conditionally independent sensory cues. To test this result in the owl, we must first determine whether IPD and ILD are conditionally independent cues. Dependence between IPD and ILD can be due to environmental variability or noise in neural computation.

We first examined whether IPD and ILD are conditionally independent when considering the variability due to the presence of environmental noise induced by concurrent sounds. For each target direction, sounds were filtered by owl HRTFs at the target direction and other directions of a second source, following the same approach used to measure the individual variability of IPD and ILD. Figure 4 shows kernel density estimates of the joint distribution of IPD and ILD at three target directions, along with density estimates assuming conditional independence. The examples are at the first, second, and third quartiles of the Kullback–Leibler divergence between the joint density and the conditional-independence approximation (Fig. 4a–c). The close match between the conditional-independence approximation and the joint density is also seen in the large fractional energy carried by the first singular value of the singular value decomposition of the joint density, which is a measure of how accurately the joint density can be approximated by a product of functions of IPD and ILD alone (median = 0.98, interquartile range = 0.038). Additionally, we found low correlation between IPD and ILD variability over different concurrent noise directions (mean absolute correlation = 0.15, s.d. = 0.12, $p < 10^{-3}$). These results are consistent with the environmental cues IPD and ILD being approximately conditionally independent at the input. We then tested whether IPD and ILD remained conditionally independent downstream in the sound localization pathway.

It is expected that IPD and ILD information remain conditionally independent within the sound localization system because IPD and ILD are processed in separate pathways (Takahashi et al. 1984). To test whether IPD and ILD cues remained conditionally independent down to ICx space specific neurons, we examined *in vivo* intracellular responses of these neurons to IPD and ILD. Specifically, we tested whether

Fig. 4 Conditional independence of IPD and ILD cues. **a-c** Environmental variability: (*top*) Kernel density estimates of the joint distribution of IPD and ILD induced by the presence of concurrent sources for three target sound directions; (*bottom*) Kernel density estimates assuming conditional independence for the same directions. The examples are at the first (**a**), second (**b**), and third (**c**) quartiles of the Kullback–Leibler divergence between the joint density and the independent approximation. **d** Three representative examples of normalized recordings of intracellular ICx membrane potential responses to broadband noise with varying ILD at a fixed IPD. Responses to correlated noise (dashed grey, BC = 1) match the responses when binaural correlation is 0 and IPD is undefined (solid black, BC = 0)

the ILD likelihood depends on IPD, i.e. whether $p_{\mathbf{s}_{ILD}|\mathbf{s}_{IPD},\Theta,\Phi}(ILD|IPD,\theta,\phi) = p_{\mathbf{s}_{ILD}|\Theta,\Phi}(ILD|\theta,\phi)$. In the barn owl's auditory system, IPD and ILD are initially processed in parallel pathways before converging in the inferior colliculus (Moiseff and Konishi 1983; Takahashi et al. 1984). To test for conditional independence of IPD and ILD, we examined membrane potential responses of ICx neurons to ILD for different conditions on IPD. Varying binaural correlation changes the reliability of the IPD cue (Jeffress et al. 1962; Albeck and Konishi 1995; Saberi et al. 1998; Egnor 2001; Peña and

Konishi 2004). When the sound signals at the left and right ears are uncorrelated, the IPD is not defined because the sounds at the left and right ears are not delayed versions of each other and the responses of ICx neurons to ILD reflect the probability distribution $p_{\mathbf{s}_{ILD}|\Theta,\Phi}(ILD|\theta,\phi)$. We therefore used responses to ILD measured with uncorrelated sounds as an estimate of the probability distribution of ILD that does not depend on IPD: $p_{\mathbf{s}_{ILD}|\Theta,\Phi}(ILD|\theta,\phi)$. Conversely, we used responses to ILD with correlated sounds as a measure of the probability distribution of ILD given IPD $p_{\mathbf{s}_{ILD}|\mathbf{s}_{IPD},\Theta,\Phi}(ILD|IPD,\theta,\phi)$. If

ILD and IPD are conditionally independent, ILD should have the same distribution in these cases and the responses should be highly correlated. In fact, the membrane potential responses of ICx neurons to ILD for uncorrelated sounds were highly correlated with the responses to ILD when IPD is present in the stimulus ($n = 12$; median correlation 0.97, interquartile range = 0.045; Fig. 4d). Additionally, if ILD and IPD are conditionally independent, then ILD should have the same distribution for any value of IPD and, similarly, IPD should have the same distribution for any value of ILD. This would predict that ILD tuning is invariant to changes in IPD and vice versa. Previous analyses have established that this holds in the owl's ICx (Peña and Konishi 2001, 2004).

Thus the ILD cue is transmitted to the midbrain independently from IPD, consistent with the claim that the cues IPD and ILD, as processed and encoded by the owl's sound localization pathway, are conditionally independent (Egnor 2001; Peña and Konishi 2004). Together, the measures of correlated variability at the input signals and neural responses suggest that IPD and ILD cues are approximately conditionally independent.

14 Sound localization in azimuth and elevation predicted by a Bayesian model

Because ILD and IPD cues are approximately conditionally independent, we modeled the noise corrupting these cues as Gaussian and circular Gaussian random vectors, respectively, with mutually independent components. With this noise model, the likelihood function is a product of IPD and ILD likelihoods (Eqs. 2–4). The noise variances are a sum of a frequency-dependent component determined by the environmental variability of IPD and ILD that we described above and a frequency-independent component modeling the neural noise that was fit to the owl's behavior (Materials and Methods).

The prior in the Bayesian model of the owl's sound localization predicts the bias for directions near the center of gaze (Edut and Eilam 2004; Fischer and Peña 2011). We modeled the prior as a product of two functions, one of azimuth and another of elevation. Based on a previous study of localization in azimuth (Fischer and Peña 2011), the azimuthal component emphasizes directions in the front. The density in azimuth was modeled as a Laplace density with zero mean and unknown variance (Eq. 6). In contrast to azimuth, where directions to the left and right of the owl have the same behavioral significance, directions above and below the owl may have different significance. Perched owls will spend time localizing prey from the perched position before flying to capture it (Ohayon et al. 2006; Fux and Eilam 2009a, b). Thus, while perched, directions below the owl may be more likely directions for prey. Therefore, we modeled the elevation component of the prior as a piecewise combination of two Gaussians

with unknown variances and common mean to allow directions above and below the owl to be treated differently (Eq. 5). The prior thus has four unknown parameters corresponding to the variance in azimuth, the two width parameters in elevation, and the center in elevation.

After fitting the parameters of the model to the behavioral data, the performance of the Bayesian estimator matched the owl's localization behavior (Fig. 5). The Bayesian model underestimated source directions in azimuth and elevation, similarly to the owl's behavior (Knudsen et al. 1979). The root-mean-square error between the average estimates reported for two owls (Knudsen et al. 1979) and the Bayesian estimate was 2.4° in azimuth and 7.9° in elevation. The differences between the Bayesian model and the owl's behavior lie within the angular discrimination limits of the owl in azimuth and elevation (Bala et al. 2007). Therefore, a Bayesian model relying on conditionally independent IPD and ILD cues describes the owl's sound localization behavior in two dimensions.

15 Neural implementation of Bayesian sound localization in azimuth and elevation

We tested the predictions of the neural implementation of optimal cue combination using a neural model of the owl's sound localization pathway (Materials and Methods). In this model, the tuning of the neurons to IPD and ILD was determined by the form of the likelihood function (Fig. 6a), while the preferred directions across the population were drawn from the prior that was found by fitting the Bayesian model to the owl's behavior (Fig. 6b–d). The source direction in azimuth and elevation was estimated by a population vector (PV; Eq. 11), which has been shown to accurately predict the owl's localization behavior in azimuth (Fischer and Peña 2011). With this network, a PV estimate of the source direction in azimuth and elevation matched both the Bayesian estimate and the owl's behavior (Fig. 5). This result is expected, based on the mathematical argument that the PV will accurately approximate the Bayesian estimate when the preferred directions are drawn from the prior and the population response is proportional to the likelihood function (Fischer and Peña 2011). We then compared the neural representation of IPD and ILD in the model network to that found in the owl's midbrain.

The modeled multiplicative responses of ICx neurons encoded the product of IPD- and ILD-based likelihoods: $f_n(\text{IPD}, \text{ILD}) \propto p_{s_{\text{IPD}}|\Theta, \Phi}(\text{IPD}|\theta_n, \phi_n) p_{s_{\text{ILD}}|\Theta, \Phi}(\text{ILD}|\theta_n, \phi_n)$. The resulting multiplicative responses to IPD and ILD are consistent with the multiplicative responses reported previously in the owl's midbrain (Peña and Konishi 2001, 2004; Fischer et al. 2007; Takahashi 2010) (Fig. 6a). This shows that the multiplicative responses observed in the owl's midbrain can support optimal Bayesian cue combination.

Fig. 5 Bayesian model performance in two-dimensional localization. The Bayesian model (Bayes) and population vector (PV) match the performance of two owls (Knudsen et al. 1979) in elevation (**a**) and azimuth (**b**). *Error bars* represent standard deviations over trials
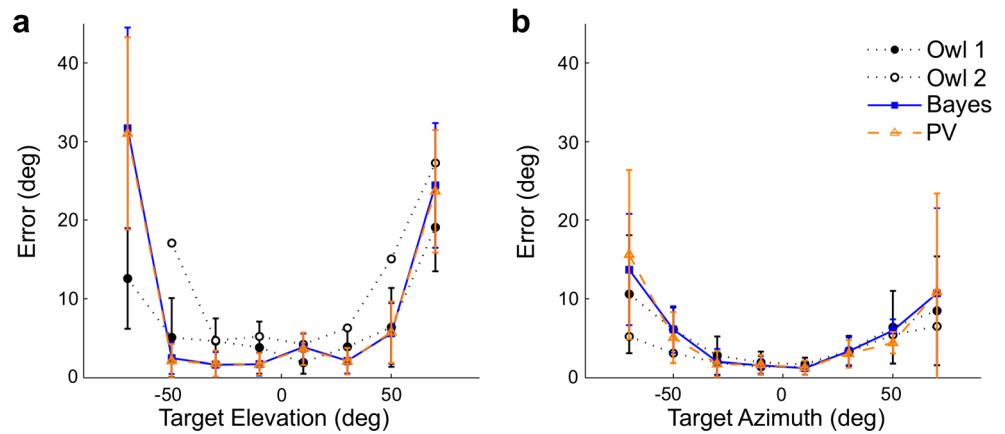
The non-uniform prior distribution of direction in the Bayesian model and the distribution of preferred directions in the owl's OT both emphasize directions near the center of gaze (Fig. 6b) (Knudsen et al. 1977; Knudsen 1982). As shown before (Fischer and Peña 2011), the density of preferred directions in azimuth predicted by the Bayesian model was consistent with the measured density in OT (Fig. 6c). However, the distribution of preferred directions in elevation was wider for the Bayesian model than the estimated density based on the shape of the OT auditory space map (Fig. 6d) (Knudsen 1982). The difference could be due to our method to estimate the prior from the shape of the space map. There are fewer data points in the periphery of the map for elevation than for azimuth, and this can cause more imprecision in the estimated

shape of the distribution in elevation than in azimuth. Yet, for both the model and the owl's space map, the highest density of directions in elevation was located below the horizontal plane. Therefore, the distribution of preferred elevations in OT is non-uniform and emphasizes directions below the owl, as seen in the prior distribution of the Bayesian model.

16 Nonlinear versus linear cue combination

To verify that multiplicative cue combination is necessary for optimal localization performance of the PV decoder, we tested this decoder on a population where model neurons combined IPD and ILD inputs additively (Fig. 7). Whether cue

Fig. 6 Neural implementation of the Bayesian model. **a** Model responses (*top*) are multiplicative, as seen in the owl's ICx (bottom, data from (Takahashi 2010), reproduced with permission). The ILD-alone and IPD-alone plots are responses when only ILD or IPD is allowed to vary. **b** The Bayesian prior in two dimensions. **c**, **d** The prior (*red*) matches the distribution of preferred directions in azimuth **c** in the owl's OT (*dashed blue*; (Knudsen 1982)). The prior in elevation is wider than the distribution of preferred elevations in OT, but both emphasize directions slightly below the center of gaze, which is indicated by the vertical line (**d**)
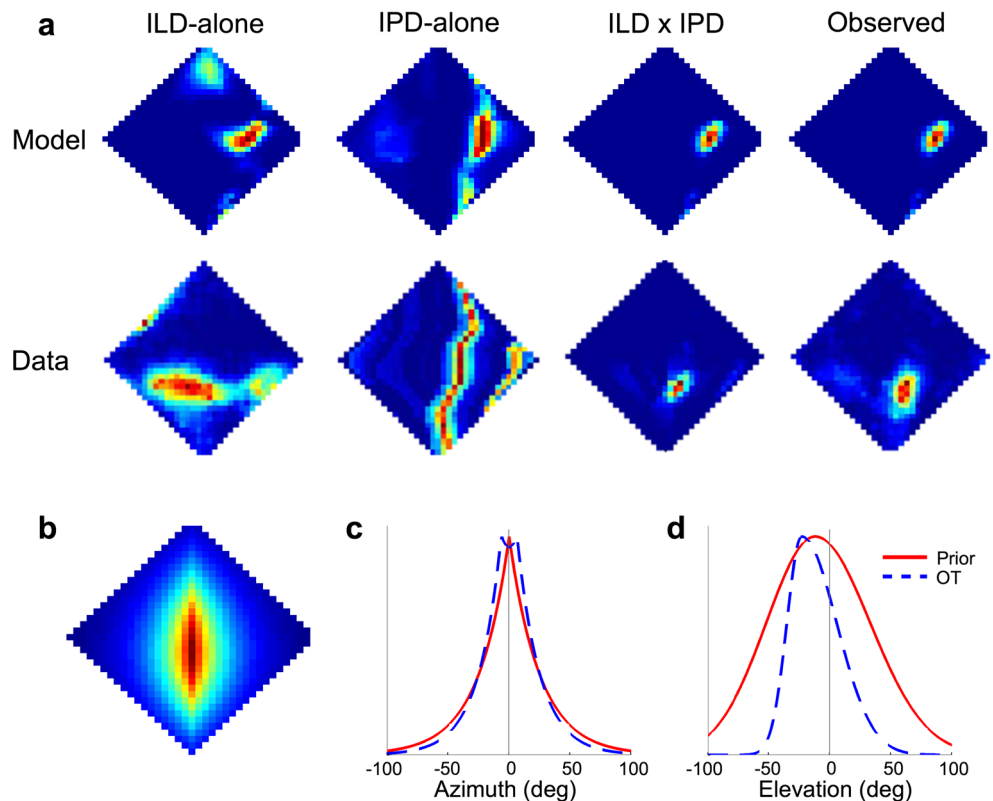
combination is linear or multiplicative greatly influences how sound source directions are represented in the auditory space map (Fig. 7a–h). Linear cue combination predicts increased activity at any place in the map where either preferred azimuth or elevation is consistent with the sensory IPD and ILD. By contrast, multiplicative cue combination predicts neural activity only at points in the map where both the preferred azimuth and elevation are consistent with the sensory IPD and ILD. With linear cue combination, the error increased dramatically in the periphery in both azimuth and elevation (Fig. 7i, j). This occurs because with additive responses, neurons with preferred directions near the center of gaze also respond when the source direction is in the periphery (Fig. 7g, h). Thus, additive responses cause source directions in the periphery to be confused with directions at the center of gaze, leading to unrealistically large errors in localization. These results show that for the PV decoder to perform optimally, cue combination must be multiplicative.

17 Discussion

This study specified the role that nonlinear operations can play in optimal cue combination. We determined that conditionally independent sensory cues combined multiplicatively can support optimal estimate of the value of an unknown stimulus. This result predicts the robust multiplicative tuning to IPD and ILD that is observed in the owl's midbrain and provides further meaning to the nonlinear integration of sensory cues within and across sensory modalities.

We have previously shown that a Bayesian model describes the owl's localization behavior in azimuth for stationary (Fischer and Peña 2011) and moving (Cox and Fischer 2015) sources. The current work is the first analysis of optimal cue combination in this particular Bayesian framework. Here we extended previous analyses to show that the Bayesian model describes behavior in both azimuth and elevation. We also derived conditions for the PV to perform Bayesian cue combination. Therefore, the Bayesian framework explains localization in azimuth of stationary sources using ITD (Fischer and Peña 2011), localization in azimuth of moving sources using ITD (Cox and Fischer 2015), and, here, localization in azimuth and elevation using IPD and ILD.

The primary assumptions of this modeling framework are that the prior is represented in the distribution of preferred directions, that the likelihood is represented in the pattern of activity across the population and that the population is readout through a PV. There is clear experimental evidence for a non-

Fig. 7 Localization with linear cue combination. Two examples (one in each row) of model ICx neuron responses where only ILD (a, b) or IPD (c, d) was let to vary with space for computing the responses. A nonlinear (multiplicative) combination of the ILD-alone and IPD-alone responses (IPD × ILD; e, f) produces a response only at directions that are consistent with both cues. A linear (additive) combination (IPD + ILD; g, h) produces responses at direction that are consistent with either cue. (i, j) The population vector fails to match the owl's performance (Knudsen et al. 1979) in elevation (i) and azimuth (j) when model neurons combine ILD and IPD inputs linearly
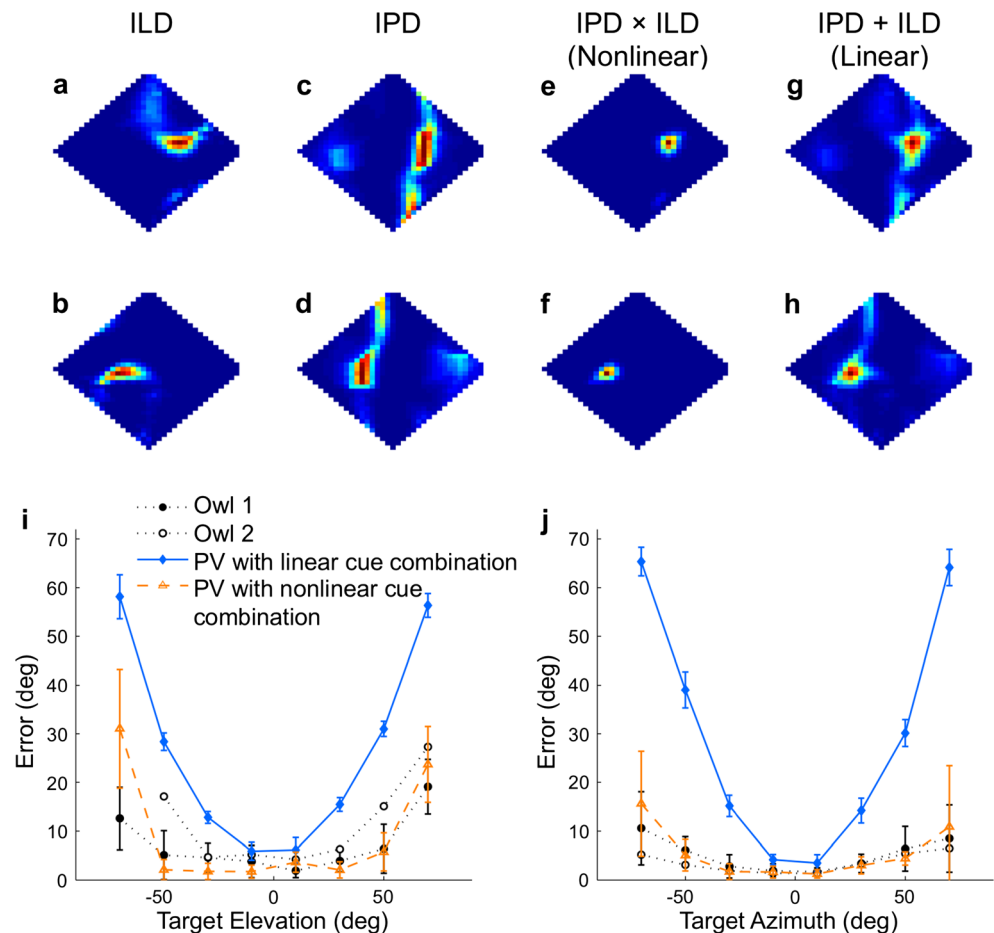
uniform distribution of preferred directions in the midbrain space-map that matches the prior distribution in the model (Knudsen 1982; Fischer and Peña 2011). We have also found experimental support for the likelihood being represented in the pattern of activity across the population. In particular, the selectivity of the neurons decreases when sensory noise increases, so that activity is spread over more of the map (Cazettes et al. 2016). Moreover, the selectivity of neurons is lower in the periphery of the space map, where IPD is a less reliable cue. The present work highlights that the well-known multiplicative tuning of space specific neurons (Peña and Konishi 2001; Fischer et al. 2007) allows the space-map responses to match the nonlinear combination of IPD and ILD joint likelihood.

While our analysis leads to the prediction that optimal cue combination is nonlinear, some other approaches suggest that optimal cue combination is linear (Jazayeri and Movshon 2006; Ma et al. 2006; Beck et al. 2007). The differences are related to the way probabilities are assumed to be represented in neural populations. How probabilities are represented may depend on the type of inference problem that involves the neural population. Here, we describe optimal cue combination for a Bayesian estimation problem where the goal is to estimate a continuous variable (sound source direction) from the posterior distribution. Alternatively, others have studied cue combination in the context of a discrimination task where the goal is to choose between discrete stimulus conditions (Fetsch et al. 2011). An optimal approach to solving the discrimination task is to compare the log-likelihood ratio to a threshold (Van Trees 2004). This operation may be computed in a neural circuit using linear operations on neural responses (Gold and Shadlen 2001) and could generalize to the case of cue combination. Our prediction of nonlinear cue combination addresses the case of producing an optimal estimate of a continuous variable using a PV; we would not arrive at the prediction of nonlinear cue combination if the goal was to represent the log-likelihood ratio in order to perform a discrimination task. The different predictions for cue-combining neurons may also be related to different assumptions for how probabilities are represented in neural populations, even for the same task. To specify the differences we will consider an inference problem where the goal is to estimate an environmental variable X (e.g. object location) based on two sensory cues C_1 and C_2 (e.g. auditory and visual input) that are initially encoded in two separate populations with responses r_1 and r_2 and then combined in a population with response r_3. In a probabilistic population code (PPC), it is assumed that the brain produces an estimate of X from the posterior distribution $p_{X|r_3}(X|r_3)$ (Ma et al. 2006; Beck et al. 2007). It has been shown that if the neurons have Poisson-like variability, then the optimal cue combination strategy is for r_3 to be a linear combination of the input activities r_1 and r_2 (Ma et al. 2006; Beck et al. 2007). This prediction arises from the assumption that the brain is drawing inferences from a posterior distribution $p_{X|r_3}(X|r_3)$ that models the neural variability in the population r_3. However,

the Poisson-like variability describes the neural variability for repeated presentations of the same stimulus and does not include the statistical relationship between the environmental cause X and the sensory stimulus, which is a major component of the overall statistics of sensory information. Also, a distribution over the population response r_3 is a very high-dimensional distribution, where the dimensionality matches the number of neurons in the population. Such a high-dimensional distribution may be difficult to learn. A similar alternative model assumes that inferences are computed using a log-likelihood function $\ln\left(p_{r_1,r_2|X}(r_1,r_2|X)\right) = \ln\left(p_{r_1|X}(r_1|X)p_{r_2|X}(r_2|X)\right) = \ln\left(p_{r_1|X}(r_1|X)\right) + \ln\left(p_{r_2|X}(r_2|X)\right)$ (Gold and Shadlen 2001; Jazayeri and Movshon 2006). Here, the logarithm transforms the nonlinear problem to adding log-likelihood functions for the input populations. Extending the model of (Jazayeri and Movshon 2006) for the representation of a log-likelihood function to cue combination would allow for optimal cue combination using linear operations. This model also assumes that the brain represents high-dimensional probability distributions that only describe the Poisson-like neural variability. Our approach is based on an alternative framework that assumes that the population r_3 represents the low-dimensional distribution $p_{X|C_1,C_2}(X|C_1,C_2)$ that describes the relationship between the environmental variables and the sensory cues. In this framework, the neural Poisson-like variability for repeated presentations of the same stimulus is treated as noise. While this is suboptimal, it simplifies the probability distribution that the brain must learn while representing the statistical relationship between the environment and the cues, which is the central component of the perceptual inference problem. Furthermore, for large populations, inferences made using the low-dimensional distribution $p_{X|C_1,C_2}(X|C_1,C_2)$ may closely approximate inferences made using the high-dimensional distribution $p_{X|r_1,r_2}(X|r_1,r_2)$ (Cazettes et al. 2016).

Nonlinear combinations have also been used in models of marginalization. In the PPC framework (Beck et al. 2011), the sound localization problem could be viewed as a marginalization problem, where the goal is to make inferences from the posterior $p_{\Theta,\Phi|r_3}(\theta,\phi|r_3)$ as:

$$p_{\Theta,\Phi|r_3}(\theta,\phi|r_3) = \int p_{\Theta,\Phi|s}(\theta,\phi|\text{IPD},\text{ILD})p_{s|\Theta,\Phi}(\text{IPD},\text{ILD}|r_1,r_2)d\text{IPD}d\text{ILD}$$

$$= \int p_{\Theta,\Phi|s}(\theta,\phi|\text{IPD},\text{ILD})p_{s_{\text{IPD}}|r_1}(\text{IPD}|r_1)p_{s_{\text{ILD}}|r_2}(\text{ILD}|r_2)d\text{IPD}d\text{ILD}.$$

This may be accomplished using a basis function network with parameters that are optimized to preserve information (Beck et al. 2011). The basis function network may use multiplicative responses in the hidden layer, although a different form of nonlinearity may possibly provide better performance. In this framework, some form of nonlinear cue combination in the basis function layer would be required, but the analysis does not predict that multiplicative responses are optimal for

cue combination in general. Experimental testing of the neural basis of marginalization will be necessary to prove the plausibility of these theories.

We propose that the responses of cue-combining neurons are given by a product of functions of the separate cues. Therefore, scaling the amplitude of one input scales the overall response. For cue reliability to influence behavior, a decrease in reliability must cause an increase in tuning widths for that cue. It has been shown that changing the reliability of IPD causes IPD tuning curves to widen, consistent with the widening of the likelihood function as reliability changes (Cazettes et al. 2016). This is distinct from a reweighting of inputs in linear neural responses that is predicted by the PPC model (Fetsch et al. 2011). The prediction of our optimal encoding model is that when the reliability of the cues change, the optimal form of cue combination remains multiplicative. For example, changing the reliability of IPD will affect the IPD-based likelihood $p_{\mathbf{s}_{\mathrm{IPD}}|\Theta,\Phi}(\mathrm{IPD}|\theta_n, \phi_n)$, but it will not change the ILD-based likelihood $p_{\mathbf{s}_{\mathrm{ILD}}|\Theta,\Phi}(\mathrm{ILD}|\theta_n, \phi_n)$, nor will it change the prediction that neural responses should be multiplicative. Intracellular *in vivo* recordings of responses to ITD and ILD in ICx showed that multiplicative cue combination is in fact robust to changes in cue reliability (Peña and Konishi 2004), which is consistent with the optimal cue combination model. These results suggest that ICx responses are consistent with the product of IPD- and ILD-based likelihoods for conditionally independent cues at different levels of reliability for IPD.

Linear and multiplicative cue combinations make different predictions for how sound source directions are represented in the midbrain auditory space and how source direction can be optimally decoded. Multiplicative cue combination predicts that activity is more localized over the map, compared to the activity predicted by linear cue combination (Fig. 7). The multiplicative model is therefore more energy efficient than the linear model by reducing the number of neurons that spike in response to a stimulus (Niven and Laughlin 2008). The restricted activity in regions of the map where both preferred IPD and ILD match the sensory input allows the sound source direction to be estimated optimally from the population responses with a PV. By contrast, if a linear combination rule is used, then an operation to detect the maximum is typically used to decode direction, which is a highly noisy mechanism (Simoncelli 2009).

There are several possible mechanisms for generating multiplicative neural responses (Koch 2004). It is possible to approximate multiplication by addition and thresholding (Fischer et al. 2009). Also, network mechanisms that depend on stimulus saliency and descending inputs can select one region of activity in the population response (Mysore and Knudsen 2012, 2013) creating a multiplicative response. The evidence is consistent with the hypothesis that multiplicative tuning in the owl's midbrain is generated in stages of linear-threshold neurons that first produce nonlinear tuning to IPD and ILD within frequency channels and then produce

nonlinear tuning across frequency (Fischer et al. 2009). Further work is required to determine how nonlinear tuning first arises in the midbrain and the role that recurrent connections play in shaping the responses.

In summary, we provided a theoretical justification for conditions under which optimal cue combination must be nonlinear and showed that these conditions are met in the owl's sound localization system. These results expand the functional implication of the robust multiplicative tuning to IPD and ILD that is observed in the owl's midbrain (Peña and Konishi 2001, 2004) by showing that multiplicative responses can allow for neurons to represent environmental statistics of multiple conditionally independent cues. This finding may apply to other cases of nonlinear cue integration within and across sensory modalities (Stein and Stanford 2008; Xu et al. 2012).

Compliance with ethical standards

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

Albeck, Y., & Konishi, M. (1995). Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals. *Journal of Neurophysiology, 74*, 1689–1700.

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology, 19*, 452–458.

Bala, A. D. S., Spitzer, M. W., & Takahashi, T. T. (2007). Auditory spatial acuity approximates the resolving power of space-specific neurons. *PLoS ONE, 2*, e675.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, 20*, 1391–1397.

Beck, J., Ma, W. J., Latham, P. E., & Pouget, A. (2007). Probabilistic population codes and the exponential family of distributions. *Progress in Brain Research, 165*, 509–519.

Beck, J. M., Latham, P. E., & Pouget, A. (2011). Marginalization in neural circuits with divisive normalization. *Journal of Neuroscience, 31*, 15310–15319.

Brainard, M. S., & Knudsen, E. I. (1993). Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl. *Journal of Neuroscience, 13*, 4589–4608.

Brainard, M. S., Knudsen, E. I., & Esterly, S. D. (1992). Neural derivation of sound source location: resolution of spatial ambiguities in binaural cues. *Journal of the Acoustical Society of America, 91*, 1015–1027.

Bregman AS (1994) Auditory scene analysis: The perceptual organization of sound. Cambridge, MA.: MIT press.

Carr, C. E., & Konishi, M. (1990). A circuit for detection of interaural time differences in the brain stem of the barn owl. *Journal of Neuroscience, 10,* 3227–3246.

Cazettes, F., Fischer, B. J., & Pena, J. L. (2014). Spatial cue reliability drives frequency tuning in the barn Owl's midbrain. *eLife, 3,* e04854.

Cazettes, F., Fischer, B. J., & Peña, J. L. (2016). Cue reliability represented in the shape of tuning curves in the owl's sound localization system. *Journal of Neuroscience, 36,* 2101–2110.

Christianson, G. B., & Peña, J. L. (2006). Noise reduction of coincidence detector output by the inferior colliculus of the barn owl. *Journal of Neuroscience, 26,* 5948–5954.

Cox, W., & Fischer, B. J. (2015). Optimal prediction of moving sound source direction in the owl. *PLoS Computational Biology, 11,* e1004360.

Edut, S., & Eilam, D. (2004). Protean behavior under barn-owl attack: voles alternate between freezing and fleeing and spiny mice flee in alternating patterns. *Behavioural Brain Research, 155,* 207–216.

Egnor, R. S. (2001). Effects of binaural decorrelation on neural and behavioral processing of interaural level differences in the barn owl (Tyto alba). *Journal of Comparative Physiology A, 187,* 589–595.

Eliasmith C., & Anderson C.H. (2004). Neural engineering: Computation, representation, and dynamics in neurobiological systems. Cambridge, MA: MIT Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience, 15,* 146–154.

Fischer, B. J., & Konishi, M. (2008). Variability reduction in interaural time difference tuning in the barn owl. *Journal of Neurophysiology, 100,* 708–715.

Fischer, B. J., & Peña, J. L. (2011). Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience, 14,* 1061–1066.

Fischer, B. J., Peña, J. L., & Konishi, M. (2007). Emergence of multiplicative auditory responses in the midbrain of the barn owl. *Journal of Neurophysiology, 98,* 1181–1193.

Fischer, B. J., Anderson, C. H., & Peña, J. L. (2009). Multiplicative auditory spatial receptive fields created by a hierarchy of population codes. *PLoS One, 4,* e8015.

Fux, M., & Eilam, D. (2009a). How barn owls (Tyto alba) visually follow moving voles (Microtus socialis) before attacking them. *Physiology and Behavior, 98,* 359–366.

Fux, M., & Eilam, D. (2009b). The trigger for barn owl (Tyto alba) attack is the onset of stopping or progressing of the prey. *Behavioural Processes, 81,* 140–143.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14,* 926–932.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science, 5,* 10–16.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience, 11,* 1201–1210.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision, 4,* 1.

Hollensteiner, K. J., Pieper, F., Engler, G., König, P., & Engel, A. K. (2015). Crossmodal integration improves sensory detection thresholds in the ferret. *PLoS One, 10,* e0124952.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research, 39,* 3621–3629.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9,* 690–696.

Jeffress, L. A., Blodgett, H. C., & Deatherage, B. H. (1962). Effect of interaural correlation on the precision of centering a noise. *Journal of the Acoustical Society of America, 34,* 1122–1123.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America, 68,* 1115–1122.

Keller, C. H., Hartung, K., & Takahashi, T. T. (1998). Head-related transfer functions of the barn owl: measurement and neural responses. *Hearing Research, 118,* 13–34.

Kita, H. & Armstrong, W. (1991). A biotin-containing compound N-(2-aminoethyl)biotinamide for intracellular labeling and neuronal tracing studies: comparison with biocytin. *Journal of Neuroscience Methods, 37,* 141–150.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27,* 712–719.

Knudsen, E. I. (1982). Auditory and visual maps of space in the optic tectum of the owl. *Journal of Neuroscience, 2,* 1177–1194.

Knudsen, E. I. (1983). Subdivisions of the inferior colliculus in the barn owl (Tyto alba). *Journal of Comparative Neurology, 218,* 174–186.

Knudsen, E. I., & Konishi, M. (1978). A neural map of auditory space in the owl. *Science, 200,* 795–797.

Knudsen, E. I., & Konishi, M. (1979). Mechanisms of sound localization in the barn owl (Tyto alba). *Journal of Comparative Physiology, 133,* 13–21.

Knudsen, E. I., Konishi, M., & Pettigrew, J. D. (1977). Receptive fields of auditory neurons in the owl. *Science, 198,* 1278–1280.

Knudsen, E. I., Blasdel, G. G., & Konishi, M. (1979). Sound localization by the barn owl (Tyto alba) measured with the search coil technique. *Journal of Comparative Physiology, 133,* 1–11.

Koch, C. (2004). *Biophysics of computation: information processing in single neurons*. USA: Oxford University Press.

Konishi, M. (1993). Neuroethology of sound localization in the owl. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology, 173,* 3–7.

Köppl, C. (1997a). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. *Journal of Neuroscience, 17,* 3312–3321.

Köppl, C. (1997b). Frequency tuning and spontaneous activity in the auditory nerve and cochlear nucleus magnocellularis of the barn owl Tyto alba. *Journal of Neurophysiology, 77,* 364–377.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22,* 79–86.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A, 18,* 2307.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research, 35,* 389–412.

Landy MS, Banks MS, Knill DC (2011) Ideal-observer models of cue integration. Sensory Cue Integration: 5–29.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9,* 1432–1438.

Manley, G. A., Koppl, C., & Konishi, M. (1988). A neural map of interaural intensity differences in the brain stem of the barn owl. *Journal of Neuroscience, 8,* 2665–2676.

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology, 42,* 135–159.

Mogdans, J., & Knudsen, E. I. (1994). Representation of interaural level difference in the VLVp, the first site of binaural comparison in the barn owl's auditory system. *Hearing Research, 74,* 148–164.

Moiseff, A. (1989). Bi-coordinate sound localization by the barn owl. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology, 164,* 637–644.

Moiseff, A., & Konishi, M. (1983). Binaural characteristics of units in the owl's brainstem auditory pathway: precursors of restricted spatial receptive fields. *Journal of Neuroscience, 3*, 2553–2562.

Mysore, S. P., & Knudsen, E. I. (2012). Reciprocal inhibition of inhibition: a circuit motif for flexible categorization in stimulus selection. *Neuron, 73*, 193–205.

Mysore, S. P., & Knudsen, E. I. (2013). A shared inhibitory circuit for both exogenous and endogenous control of stimulus selection. *Nature Neuroscience, 16*, 473–478.

Niven, J. E., & Laughlin, S. B. (2008). Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology, 211*, 1792–1804.

Nix, J., & Hohmann, V. (2006). Sound source localization in real sound fields based on empirical statistics of interaural parametersa). *Journal of the Acoustical Society of America, 119*, 463–479.

Ohayon, S., van der Willigen, R. F., Wagner, H., Katsman, I., & Rivlin, E. (2006). On the barn owl's visual pre-attack behavior: I. Structure of head movements and motion patterns. *Journal of Comparative Physiology A, 192*, 927–940.

Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience, 14*, 775–782.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*, 2451–2468.

Palmer, A. R., & Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research, 24*, 1–15.

Pecka, M., Siveke, I., Grothe, B., & Lesica, N. A. (2010). Enhancement of ITD coding within the initial stages of the auditory pathway. *Journal of Neurophysiology, 103*, 38–46.

Peña, J. L., & Konishi, M. (2000). Cellular mechanisms for resolving phase ambiguity in the owl's inferior colliculus. *Proceedings of the National Academy of Science, 97*, 11787–11792.

Peña, J. L., & Konishi, M. (2001). Auditory spatial receptive fields created by multiplication. *Science, 292*, 249–252.

Peña, J. L., & Konishi, M. (2002). From postsynaptic potentials to spikes in the genesis of auditory spatial receptive fields. *Journal of Neuroscience, 22*, 5652–5658.

Peña, J. L., & Konishi, M. (2004). Robustness of multiplicative processes in auditory spatial tuning. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 24*, 8907–8910.

Roman, N., Wang, D., & Brown, G. J. (2003). Speech segregation based on sound localization. *Journal of the Acoustical Society of America, 114*, 2236–2252.

Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B. J., & Farahbod, H. (1998). Effects of interaural decorrelation on neural and behavioral detection of spatial cues. *Neuron, 21*, 789–798.

Shi, L., & Griffiths, T.L. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems 22* (pp. 1669–1677). Cambridge, MA: MIT Press.

Simoncelli, E.P. (2009). Optimal estimation in sensory systems. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (vol. 4, pp. 525–535). Cambridge, MA: MIT Press.

Spitzer, M. W., Bala, A. D., & Takahashi, T. T. (2003). Auditory spatial discrimination by barn owls in simulated echoic conditions. *Journal of the Acoustical Society of America, 113*, 1631–1645.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Review Neuroscience, 9*, 255–266.

Takahashi, T. T. (2010). How the owl tracks its prey–II. *Journal of Experimental Biology, 213*, 3399–3408.

Takahashi, T., Moiseff, A., & Konishi, M. (1984). Time and intensity cues are processed independently in the auditory system of the owl. *Journal of Neuroscience, 4*, 1781–1786.

van Beers, R. J., Sittig, A. C., & Gon, J. J. (1999). Integration of proprioceptive and visual position-information: an experimentally supported model. *Journal of Neurophysiology, 81*, 1355–1364.

Van Trees HL (2004) Detection, estimation, and modulation theory. New York, NY: Wiley.

Wagner, H., Takahashi, T., & Konishi, M. (1987). Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 7*, 3105–3116.

Whitchurch, E. A., & Takahashi, T. T. (2006). Combined auditory and visual stimuli facilitate head saccades in the barn owl (Tyto alba). *Journal of Neurophysiology, 96*, 730–745.

Wightman, F. L. & Kistler, D. J. (1993). Sound localization. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Human psychophysics* (pp. 155–192). New York: Springer-Verlag.

Xu, J., Yu, L., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2012). Incorporating cross-modal statistics in the development and maintenance of multisensory integration. *Journal of Neuroscience, 32*, 2287–2298.