A computational model of sound localization in the barn owl

Brian J. Fischer, Charles H. Anderson

Abstract

We model the representation of auditory space in the barn owl as a posterior probability density function over location variables given cues extracted from the auditory input signals. Binaural disparity cues are introduced to the input signals through the use of head related transfer functions. The cue extraction process is based on the extraction of temporal and intensity information in parallel pathways. Inferring locations from auditory cues relies on stimulus invariance properties of the computed location cues.

1 Estimating locations

The barn owl exhibits a natural head turning behavior in response to novel sound stimuli that has been used to demonstrate the great accuracy with which the barn owl localizes sources in the horizontal and vertical dimensions [4]. The neural basis for this localization behavior as mediated by the midbrain sound localization pathway is the existence of neurons with spatially restricted receptive fields arranged in a topographic map that first arises in the external nucleus of the inferior colliculus (ICx) [5]. In the context of viewing the sound localization problem that is solved by the midbrain sound localization pathway as an estimation problem, we model the representation of auditory space in ICx as a posterior probability density function over location variables given cues extracted from the auditory input signals. By the nature of the response that the system produces, a rapid head saccade to a particular direction, we assume that the system is designed to localize single sources in the presence of an "uneventful" background.

1.1 The target

The state of the target source is defined by the signal that the source produces, $s(t)$, and the location of the source (θ, ϕ). Here θ corresponds to azimuth and ϕ corresponds to elevation where the angles are defined relative to the owl's head. We assume that the locations are random variables with ranges limited to the frontal hemisphere and that θ and ϕ are constant over a small time window. Let Ω denote the range of (θ, ϕ). We assume that the relevant signals

are broadband signals ($2 - 12$ kHz). Because the locations are defined with respect to the owl's head we assume that the prior distribution over locations is uniform, $p(\theta, \phi) = \frac{1}{\int\int_\Omega d\theta d\phi}$ for $(\theta, \phi) \in \Omega$.

1.2 Observation

The first step in the localization process is the location dependent mapping of source signals to the received pressure waveforms at the ears. The received signals are modeled as a sum of a target component and a noise component

$$\begin{aligned} r_L(t) &= s_L(t) + n_L(t) \\ &= h_{L(\theta,\phi)}(t) \star s(t) + n_L(t) \end{aligned} \tag{1}$$

and

$$\begin{aligned} r_R(t) &= s_R(t) + n_R(t) \\ &= h_{R(\theta,\phi)}(t) \star s(t) + n_R(t). \end{aligned} \tag{2}$$

The target component is a filtered version of the source signal where $h_{L(\theta,\phi)}(t)$ and $h_{R(\theta,\phi)}(t)$ are the head related impulse responses (HRIR) for the left and right ears, respectively, when the source location is (θ, ϕ) [3]. The noise signal consists of two components, background noise and measurement noise and is assumed to be Gaussian white noise bandlimited to 12 kHz with lower rms intensity than the target signal.

1.3 Representing auditory space

In our model, location information is not inferred directly from the received signals but is obtained from stimulus independent binaural location cues that are extracted from the input signals. We compute a vector of cues from the observed data, ξ_t, that contains only information about the source location. We assume that the cue vector ξ_t is a sufficient statistic for (θ, ϕ) that satisfies

$$\xi_t = \bar{\xi}(\theta, \phi) + n(t) \tag{3}$$

where $\bar{\xi}(\theta, \phi)$ is a stimulus invariant feature vector and $n(t)$ is Gaussian white noise. The assumption that ξ_t is a sufficient statistic for (θ, ϕ) means that $p(\theta, \phi|\xi_t, r_L(t), r_R(t)) = p(\theta, \phi|\xi_t)$ or simply that we have extracted from the observed data all of the information about the location of the target. Therefore we can obtain information about the source location through the posterior density

$$p(\theta, \phi|\xi_t) = \frac{p(\xi_t|\theta, \phi)p(\theta, \phi)}{\int\int p(\xi_t|\theta, \phi)p(\theta, \phi)d\theta d\phi}$$

without any loss.

The assumption that ξ_t is equal to an invariant feature vector plus Gaussian noise allows us to give a form for the likelihood, $p(\xi_t|\theta, \phi) \propto \exp(-\frac{1}{2}\|\xi_t - \bar{\xi}(\theta, \phi)\|^2_{\Sigma_n^{-1}})$ where the covariance matrix of the noise at time t is Σ_n.

With our assumption of a uniform prior, we have that the posterior density is given by

$$p(\theta, \phi|\xi_t) \propto \exp(-\frac{1}{2}\|\xi_t - \bar{\xi}(\theta, \phi)\|^2_{\Sigma_n^{-1}}). \tag{4}$$

The use of the vector of location cues instead of the observed data greatly simplifies the problem of inferring the location of the target source. This approach is also consistent with the operations that are performed in the auditory system of the barn owl. We now specify the particular cues that are used in our model.

2 Cue Extraction

2.1 Filtering and half-wave rectification

In the first stage of our model, input signals are filtered with a bank of linear band-pass filters. Following the linear filtering, input signals undergo a half-wave rectification. So, the input signals to the two ears $r_L(t)$ and $r_R(t)$ are decomposed into a set of scalar valued functions $u_L(t, \omega_k)$ and $u_R(t, \omega_k)$ defined by

$$u_L(t, \omega_k) = [f_{\omega_k} \star r_L(t)]_+ \tag{5}$$

and

$$u_R(t, \omega_k) = [f_{\omega_k} \star r_R(t)]_+ \tag{6}$$

where $f_{\omega_k}(t)$ is the linear bandpass filter for the channel with center frequency ω_k. Here we use the standard gammatone filter $f_{\omega_k}(t) = t^{\gamma-1}e^{-t/\tau_k}\cos(\omega_k t)$ with $\gamma = 4$. Initial operations of cue extraction will be performed within distinct frequency channels that are established by this filtering process.

2.2 Intensity processing

The intensity difference pathway has two stages. First, the outputs of the filter banks are integrated over time to obtain windowed intensity measures for the components of the left and right signals. Next, signals from the left and right ears are combined within each frequency channel to measure the location dependent intensity difference. We compute the intensity of the signal in each frequency band over a small time window as:

$$y_L(t, \omega_k) = \int_0^t u_L(\sigma, \omega_k) w(t - \sigma) d\sigma \tag{7}$$

and

$$y_R(t, \omega_k) = \int_0^t u_R(\sigma, \omega_k) w(t - \sigma) d\sigma \tag{8}$$

where $w(t)$ is a window. We use a simple exponential window $w(t) = e^{-t/\tau}H(t)$ where $H(t)$ is the unit step function.

The magnitude of $y_L(t,\omega_k)$ and $y_R(t,\omega_k)$ vary with both the signal intensity and the gain of the HRIR in the frequency band centered at ω_k. However, only the relative gain applied to the signals of the left and right ears by the HRIR reflects the location of the source. To compute the intensity difference between the input signals that is introduced by the HRIRs in a manner that is invariant to changes in the intensity of the source signal we compute

$$z(t,\omega_k) = \log(\frac{y_R(t,\omega_k)}{y_L(t,\omega_k)}). \tag{9}$$

2.3 Temporal processing

We use a windowed cross correlation operation to measure time differences that includes three features that model processing in the barn owl's auditory system. First, signals are passed through a saturating nonlinearity to model the saturation of the nucleus magnocellularis (NM) inputs to the nucleus laminaris (NL) [8]. We define $v_L(t,\omega_k) = F(u_L(t,\omega_k))$ and $v_R(t,\omega_k) = F(u_R(t,\omega_k))$ where $F(\cdot)$ is a saturating nonlinearity. Let $x(t,\omega_k,m)$ denote the value of the cross correlator in frequency channel k at delay index $m \in \{0,\ldots,N\}$ defined by

$$\dot{x}(t,\omega_k,m) = -\frac{x(t,\omega_k,m)}{\tau(y(t,\omega_k))} + [v_L(t-\Delta m,\omega_k)+\alpha][v_R(t-\Delta(N-m),\omega_k)+\beta] \tag{10}$$

where $\tau(y(t,\omega_k))$ is a time constant that varies with the intensity of the stimulus in the frequency channel where $y(t,\omega_k) = y_L(t,\omega_k) + y_R(t,\omega_k)$, Δ is a delay increment, and $\alpha,\beta > 0$ are included to reflect the fact that NL neurons respond to monaural stimulation [2]. The constants α andβ are chosen so that at input levels above threshold ($0-5$ dB SPL) the cross correlation term dominates. The time constant decreases as $y(t,\omega_k)$ increases so that for more intense sounds information is integrated over a smaller time window. This operation functions as a gain control and models the inhibition of NL neurons by superior olive neurons [9]. We choose Δ to satisfy $\Delta N = 200\mu s$ so that the full range of possible delays is covered.

3 Combining temporal and intensity difference signals

Space specific neurons of the ICx respond to combinations of the interaural time difference and interaural level difference that correspond to locations in space [6]. In our model the representation of space is given by a posterior density over locations conditioned on the cross correlations and intensity differences described above. To facilitate the discussion of the calculation of the posterior density over the locations we introduce notation for the cues derived from the auditory signals. Let $\mathbf{x}(t,\omega_k) = [x(t,\omega_k,0),\ldots,x(t,\omega_k,N)]/\|[x(t,\omega_k,0),\ldots,x(t,\omega_k,N)]\|$ be the normalized vector of cross correlations computed within frequency channel k. Let $\mathbf{x}(t) = [\mathbf{x}(t,\omega_1),\ldots,\mathbf{x}(t,\omega_{N_F})]$ denote the vector of cross correlations

and let $\mathbf{z}(t) = [z(t, \omega_1), \ldots, z(t, \omega_{N_F})]$ denote the vector of intensity differences where N_F is the number of frequency channels.

The transformation from cues computed from the auditory input signals to a representation of space occurs by performing inference on the cues, $\xi_t = [\mathbf{x}(t)\ \mathbf{z}(t)]$, through the posterior density

$$p(\theta, \phi|\xi_t) = \frac{1}{Z} \exp(-\frac{1}{2} \|(\mathbf{x}(t), \mathbf{z}(t)) - (\bar{\mathbf{x}}(\theta, \phi), \bar{\mathbf{z}}(\theta, \phi))\|^2_{\Sigma_n^{-1}}). \tag{11}$$

The known physiology of the barn owl places constraints on how this density can be computed. The spatial selectivity of auditory neurons in the optic tectum is consistent with a model where spatial selectivity arises from tuning to combinations of time difference and intensity difference cues within each frequency channel [1]. Also, the existence of significant side peaks in the time difference tuning of space specific neurons suggests that information is not multiplied across frequency channels [7]. Therefore, we use a model in which time difference and intensity difference information is first combined multiplicatively within frequency channels and is then summed across frequency.

So we assume that $p(\theta, \phi|\mathbf{x}(t), \mathbf{z}(t))$ is approximated by a kernel estimate of the form

$$p(\theta, \phi|\mathbf{x}(t), \mathbf{z}(t)) = \sum_k c_k K(\theta, \phi; \mathbf{x}(t, \omega_k), z(t, \omega_k)) \tag{12}$$

where each kernel is of the form

$$K(\theta, \phi; \mathbf{x}(t, \omega_k), z(t, \omega_k)) = K_x(\theta, \phi; \mathbf{x}(t, \omega_k)) K_z(\theta, \phi; z(t, \omega_k)). \tag{13}$$

We require that $\int \int K(\theta, \phi; \mathbf{x}(t, \omega_k), z(t, \omega_k)) d\theta d\phi = 1$ and $\sum_k c_k = 1$.

As an example, in figure 1 we plot the approximate posterior density at $t = 25$ ms for both a broadband noise signal and a tone at 7 kHz. In each case there is no noise added to the input signals. For the broadband noise signal there is a single peak at the true source location. For the tonal signal there is a peak at the true source location, but there are additional peaks that occur at locations where the associated binaural cues are consistent with binaural location cues of the source location.

4 Summary

We have presented a model of the representation of auditory space in ICx that is based on an estimation formulation of the sound localization problem. In our model information about the spatial location of a single target source is inferred from location dependent binaural time and intensity differences introduced by head relation impulse responses. A critical feature of the inference process is the extraction of stimulus independent cues. Stimulus independence is obtained in intensity difference pathway by taking the log of the ratio of intensity measures from the left and right and in the time difference pathway by normalizing the cross correlation vector. Cues are then combined mulitplicatively within frequency channels to obtain information about auditory space.

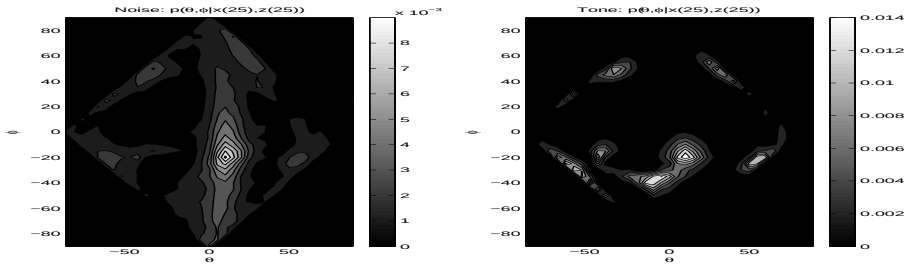

Figure 1: Left: Posterior density for a broadband noise signal located at $(20^o, -10^o)$. Right: Posterior density for a 7 kHz tonal signal located at $(20^o, -10^o)$. The densities are shown for $t = 25\ ms$.

References

[1] M.S. Brainard, E.I. Knudsen, S.D. Esterly, Neural derivation of sound source location: Resolution of spatial ambiguities in binaural cues. *J. Acoust. Soc. Am., 91(2):1015-1026, 1992*

[2] C.E. Carr, M. Konishi, A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci., 10(10):3227-3246, 1990*

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

[4] E.I. Knudsen, G.G. Blasdel, M. Konishi, Sound localization by the barn owl (Tyto alba) measured with the search coil technique. *J. Comp. Phys. A, 133:1-11, 1979*

[5] E.I. Knudsen, M. Konishi, A neural map of auditory space in the owl. *Science, 200:795-797, 1978*

[6] A. Moiseff, M. Konishi, Neuronal and behavioral sensitivity to binaural time differences in the owl. *J. Neurosci., 1(1):40-48, 1981*

[7] J.L. Peña, M. Konishi, From postsynaptic potentials to spikes in the genesis of auditory spatial receptive fields. *J. Neurosci., 22(13):5652-5658, 2002*

[8] W.E. Sullivan, M. Konishi, Segregation of stimulus phase and intensity coding in the cochlear nuclues of the barn owl. *J. Neurosci., 4(7):1787-1799, 1984*

[9] L. Yang, P. Monsivais, E.W. Rubel, The superior olivary nucleus and its influence on nucleus laminaris: A source of inhibitory feedback for coincidence detection in the avian auditory brainstem. *J. Neurosci., 19(6):2313-2325, 1999*