A mammalian inferior colliculus model for sound source separation using interaural time differences

Christian Leibold*

Fakultät für Biologie & Bernstein Center Freiburg,

Albert-Ludwigs-Universität Freiburg and

BrainLinks-BrainTools Albert-Ludwigs-Universität Freiburg

Sebastian Groß

Fakultät für Biologie, Ludwig-Maximilians Universität München

Abstract

The inferior colliculus (IC) hosts many neurons tuned to interaural time differences (ITDs) at low frequencies. ITD tuning, however, is already observed and generated in the superior olivary complex. Here, we argue that combining ITD sensitive inputs from medial superior olive (MSO) and low-frequency lateral superior olive (LSO) requires a temporally well adjusted delay of commissural fibers from LSO to IC, given the fast synaptic kinetics of IC neurons. We present a normative model of the midbrain auditory circuitry that finds an optimal commissural delay of 0.3 cycles and optimal synaptic strengths by maximizing the firing rate of IC neurons for a given ITD. The model suggests that, by varying the relative synaptic weight of MSO and LSO input, individual neurons are optimized to transmit information for any sound source in complex auditory scenes. ITD tuning of IC neurons results as a side effect and follows the statistics known from experimental recordings.

INTRODUCTION

Separation of sound sources in noisy acoustic environments is one of the most intriguing capabilities of the human auditory system and still poses a hard problem to modern artificial sound separation solutions [1]. Deficits in our ability to separate sounds in such cluttered environments usually establish years before hearing deficits are diagnosed by conventional audiological methods [2–4]. In search for the underlying neural mechanisms of sound separation, binaural hearing turned out to be of fundamental importance [5–8]. Neural mechanisms of binaural processing in the ascending auditory pathway are generally discussed in the context of the nuclei at which binaural interactions take place first [9–11]: The medial and the lateral superior olive (MSO, LSO). The majority of the available data about the neural encoding of the two main binaural cues, interaural time difference (ITD) and interaural level difference (ILD), however, stems from the higher-order inferior colliculus (IC) (e.g. [12–15]). Mostly, binaural tuning in the IC is then interpreted as a proxy of the original representation in MSO and LSO. Although there are notable exceptions that aim to explain discrepancies between IC and superior olivary binaural tuning characteristics in the light of the local circuitry [16, 17], and even suggest mechanisms of de-novo generation of ILD sensitivity [18], a normative functional approach of binaural representations in the

* christian.leiboldbiologie.uni-freiburg.de

IC based on circuit mechanisms is missing. A particular short-coming thereby arises from the sluggishness assumption [19] implying that IC activity establishes a rate code requiring long averaging windows and assuming that all fine-scale temporal structure is lost. The feed-forward excitatory synaptic kinetics in the IC, however, is still relatively fast, about 2 ms [20], and thus phase locking of the synaptic current to stimulus fine structure should still be observable up to almost about 1 kHz as confirmed by electrophysiological recordings [21–23].

In this paper, we present a normative theory of how IC neurons read the brainstem activity of MSO and LSO, assuming that both nuclei represent ITDs as two populations, each one of which only conveys the summed activity of all its neurons in a frequency band further downstream [24–27]. We argue that best ITDs of IC neurons can thereby vary over a whole hemisphere with only an adjustment of the relative synaptic weight (LSO vs. MSO) to be required. In contrast to most normative models of sound localization that maximize localization acuity [25, 28, 29], the present theory optimizes sound source separability of mixtures of sounds. It thereby naturally accounts for binaural unmasking, a well-known and long-studied effect showing the binaural improvement of detection thresholds in absence of binaural signal-to-noise differences [30, 31], and for which ITDs are the dominant cues at low frequencies [32].

Binaural unmasking has so far attracted only little attention of biologically inspired modelling [33, 34], despite highly evolved psychophysical models [7, 35–38]; for review of classical models see [39]. The present theory suggests the temporally fine-tuned interactions of afferent ITD representations at the level of the IC play a key role for the separation of low frequency sounds, combining mechanisms of binaural unmasking and sound localization.

THEORY

An auditory scene is a superposition of sounds arriving from multiple sources. Typically, only one source $S(t)$ is considered as the signal to be extracted. In this paper, the signal is assumed to have constant ITD Δt (> 0 for sounds on the right), and the same sound intensity at the two ears. Thus, sound signals arriving at the right R and left L ear are

linear combinations

$$R(t) = S\left(t + \frac{\Delta t}{2}\right) + N_R(t) \tag{1}$$

$$L(t) = S\left(t - \frac{\Delta t}{2}\right) + N_L(t) \tag{2}$$

where $N_{R/L}$ summarize the contributions of the non-signal sounds. Hence, the eardrums translate two pressure waves R and L into neural activity that encodes three unknowns S, N_R and N_L.

In Appendix A it is shown that minimizing the mean square error between a linear estimate $\tilde{S}$ of S from the two "measurements" R and L yields

$$\tilde{S}(t) = \frac{1}{3}\left[R\left(t - \frac{\Delta t}{2}\right) + L\left(t + \frac{\Delta t}{2}\right)\right] \ . \tag{3}$$

Equation 3 can be interpreted as Jeffress' delay-line model for a neuron in which the internal delay between left and right ear exactly compensates the ITD, i.e., a Jeffress-type coincidence detector neuron with the delay offset Δt optimally (in the Euclidean sense) reconstructs sound sources originating from an ITD Δt.

Jeffress-type delay lines, however, have not been found in the mammalian sound localization circuitry [40], and thus alternative solutions are of interest of how equation (3) can be implemented in ways more consistent with mammalian brainstem anatomy.

Sensitivity to low-frequency sounds has not only been reported for MSO neurons, but also for LSO neurons [44–47]. Whereas MSO neurons receive excitatory and inhibitory inputs from both ears, LSO neurons combine ipsilateral excitation with contralateral inhibition. In the simplest form one thus may model the activity of right-hemispheric MSO and left-hemispheric LSO as

$$MSO(t) = R(t) + L(t - \Lambda_m) \tag{4}$$

$$LSO(t) = L(t) - R(t - \Lambda_l) \ , \tag{5}$$

with $\Lambda_{m/l}$ denoting the cross-hemispheric latencies. MSO and LSO activity are combined in the inferior colliculus (IC) where principal excitatory pathways converge from the contralateral LSO and the ipsilateral MSO (Fig. 1 and Glendenning et al., 1992; Brückner and Rübsamen, 1995), i.e., for IC of the right hemisphere one finds,

$$\begin{aligned} IC(t) &= [MSO(t) + a\, LSO(t - \Lambda_c)]/\sqrt{1 + a^2} \\ &= [R(t) - a\, R(t - \Lambda_l - \Lambda_c) + L(t - \Lambda_m) + a\, L(t - \Lambda_c)]/\sqrt{1 + a^2} \ , \end{aligned} \tag{6}$$

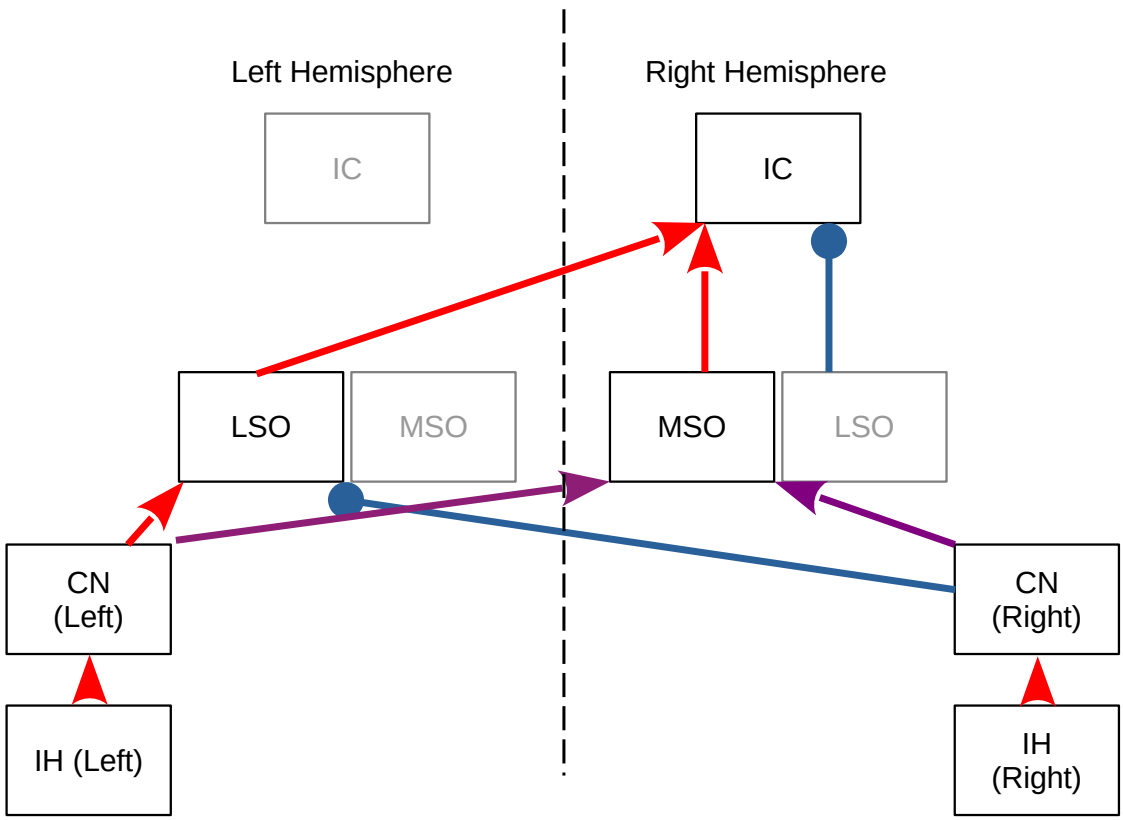

FIG. 1. Ascending binaural brainstem anatomy. The inner hair cells (IH) convert mechanical vibrations of the basilar membrane into an electrical signal that is conveyed to the ipsilateral cochlear nuclei (CN) via glutamatergic synaptic transmission (red) of spiral ganglion neurons. CN bushy cell axon innervate bilateral MSO and LSO neurons either directly via glutamatergic synapses (red) or indirectly via glycinergic synapses (blue) from the medial nucleus of the trapezoid body. Whereas for LSO neurons, the presynaptic neurotransmitters separate according to hemisphere, MSO neurons receive mixed (purple) glutamatergic and glycinergic (Kapfer and Grothe, 2002; Couchman et al., 2010; Roberts et al., 2013) synapses from each hemisphere. IC neurons receive glutamatergic synapses from ipsilateral MSO and contralateral LSO, as well as glycinergic synapses from the ipsilateral LSO. The model in this paper is devised for the right IC. Due to symmetry a left IC model can be obtained by flipping right and left ear input.

with $a > 0$ denoting the relative synaptic strength and Λ_c the relative commissural delay of the contralateral LSO input (via the stria of Held). The factor $\sqrt{1+a^2}$ normalizes the sum of squared weights to 1, removing effects on the output firing rate induced by the total synaptic drive.

So far our theory neglects frequency-specific processing, and thus, in order to increase

biological realism, equations (4-6) should be considered in a best frequency (BF)-channel dependent manner. The Fourier transform of Eq. (6) (as derived in Appendix B), would thus provide a model for the ITD tuning curve of an IC neuron in the frequency channel $\omega = 2\pi$ BF,

$$|\widehat{IC}(\omega)|^2 \;=\; |\hat{S}(\omega)|^2\,[|\gamma_R|^2 + |\gamma_L|^2 + 2\,|\gamma_R\gamma_L|\,\cos\Delta\Phi]/(1+a^2)\;, \tag{7}$$

in which $\gamma_R = 1 - a\,\exp[-\mathrm{i}(\psi_l + \psi_c)]$, $\gamma_L = \exp(-\mathrm{i}\Delta\phi)\,[\exp(-\mathrm{i}\psi_m) + a\,\exp(-\mathrm{i}\psi_c)]$, $\Delta\Phi = \arg[\gamma_R/\gamma_L]$, and the interaural phase difference $\Delta\phi = \omega\,\mathrm{ITD}$. Another advantage of the frequency domain representation is that the contralateral transmission delays $\Lambda_{l,m,c}$ can be allowed to vary with frequency ω and therefore implement any type of frequency-dependent phase $\psi(\omega) = \Lambda(\omega)\,\omega$, including characteristic phases and delays.

In this paper, we will constrain the commissural latencies presynaptic to MSO and LSO to their average experimentally reported values ($\psi_m = 0.125$ (cyc) [50]; $\psi_l = 0$ [45, 51]) and maintain ψ_c as a free parameter.

Moreover, we also allow the LSO synaptic weight a to take negative values reflecting the ipsilateral inhibitory projection [48], which would change the above formulas to $\gamma_R^{(i)} = 1 + a$, $\gamma_L^{(i)} = \exp(-\mathrm{i}\Delta\phi)[\exp(-\mathrm{i}\psi_m) - a\,\exp(-\mathrm{i}\psi_l)]$, and evidently is independent of the commissural phase delay ψ_c.

The dependence of the ITD tuning curves $|IC|^2$ (Eq. 7) on the synaptic weight a is illustrated in Fig. 2A for two different frequency channels and two commissural phases ψ_c. Increasing a amplifies the LSO contribution and can effectively shift the best ITD (position of the peak firing rate) to the left or to the right, depending on the relative commissural phase delay ψ_c. The shift in best ITD via a change in the relative synaptic weight a occurs gradually in some intervals of the a axis with larger shifts for lower best frequencies, whereas the dependence of best ITD on ψ_c appears more complex (Fig. 2B) but hints at the existence of a low-ψ_c and a high-ψ_c regime.

A systematic way to evaluate the influence of the two free parameters a and ψ_c is to search for optimal combinations. Here, a two-step optimization process was performed independently in each BF channel. First, for a given ITD and ψ_c, the optimal weight

$$a(\mathrm{ITD}, \psi_c) = \mathrm{argmax}_{a\in[-10,10]}|IC|^2$$

was chosen to maximize the firing rate of an IC cell according to equation (7). This objective deviates from the usual approach to represent auditory space using a wide range of best ITDs,

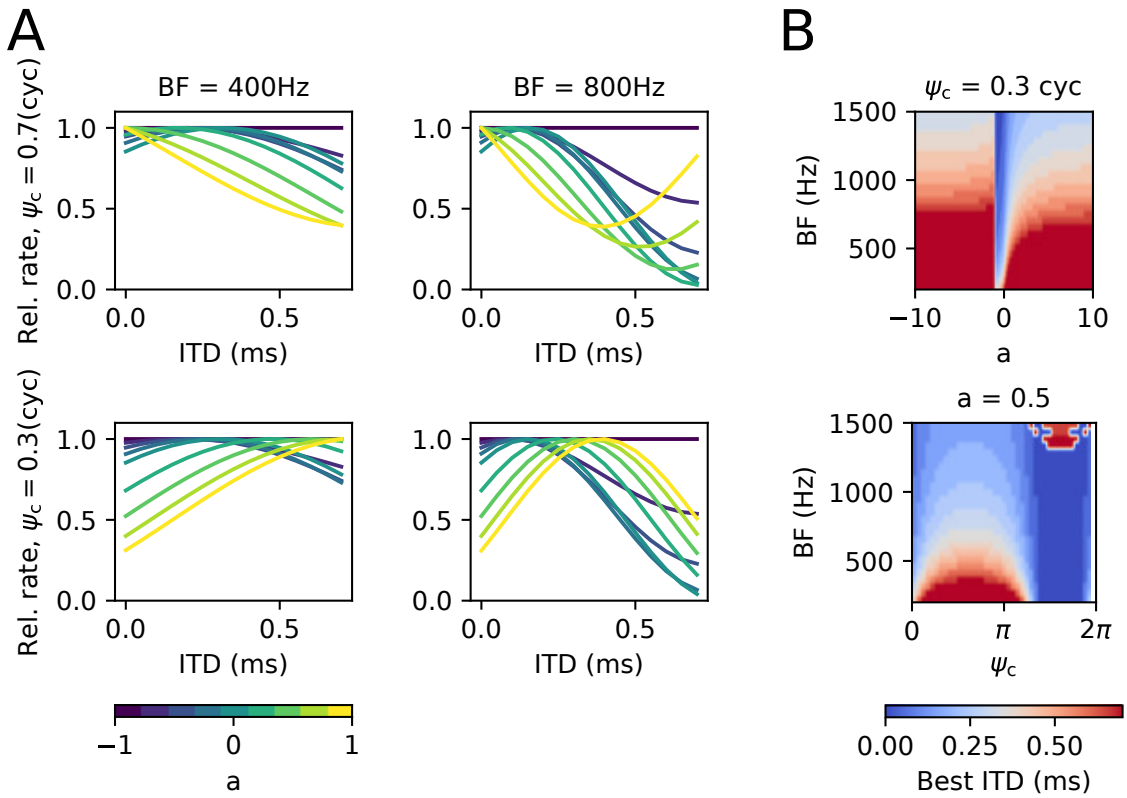

FIG. 2. Example ITD tuning curves. (A) ITD tuning curves derived from Eq. (7) are shown for four exemplary conditions and varying relative LSO weights a (colors). Top panels are obtained with a large commissural phase $\psi_c = 0.7$ cyc, bottom panels for small commissural phase $\psi_c = 0.3$ cyc. Left and right column correspond to best frequency channels of 400 and 800 Hz, respectively. (B) Best ITDs are color coded for fixed $\psi_c = 0.3$ cyc as a function of a (top) and for fixed $a = 0.5$ as a function of ψ_c.

but rather tries to transmit at maximal rate for any given signal ITD. In a second step ψ_c, was determined by minimizing the loss function

$$\mathcal{L}(\psi_c) = \int_0^{\mathrm{ITD}_{\max}} \mathrm{d}\tau \; \left(|a(\tau,\psi_c)| - \ln |a(\tau,\psi_c)| + \lambda \, \mathrm{sign}[-a(\tau,\psi_c)] \right) \; .$$

By construction, the loss function tries to minimize the mean relative LSO weight (accounting for the relatively small low frequency LSO), while the $-\ln$ term avoids pure MSO solutions with $a = 0$. The regularization $\lambda \, \mathrm{sign}[-a(\tau,\psi_c)]$ punishes the inhibitory ipsilateral pathway, which, from anatomy, is know to be much smaller in size [48]. Results in this paper were obtained with regularization factor $\lambda = 1$.

The outcomes of the optimization process are illustrated in Fig. 3 for three different

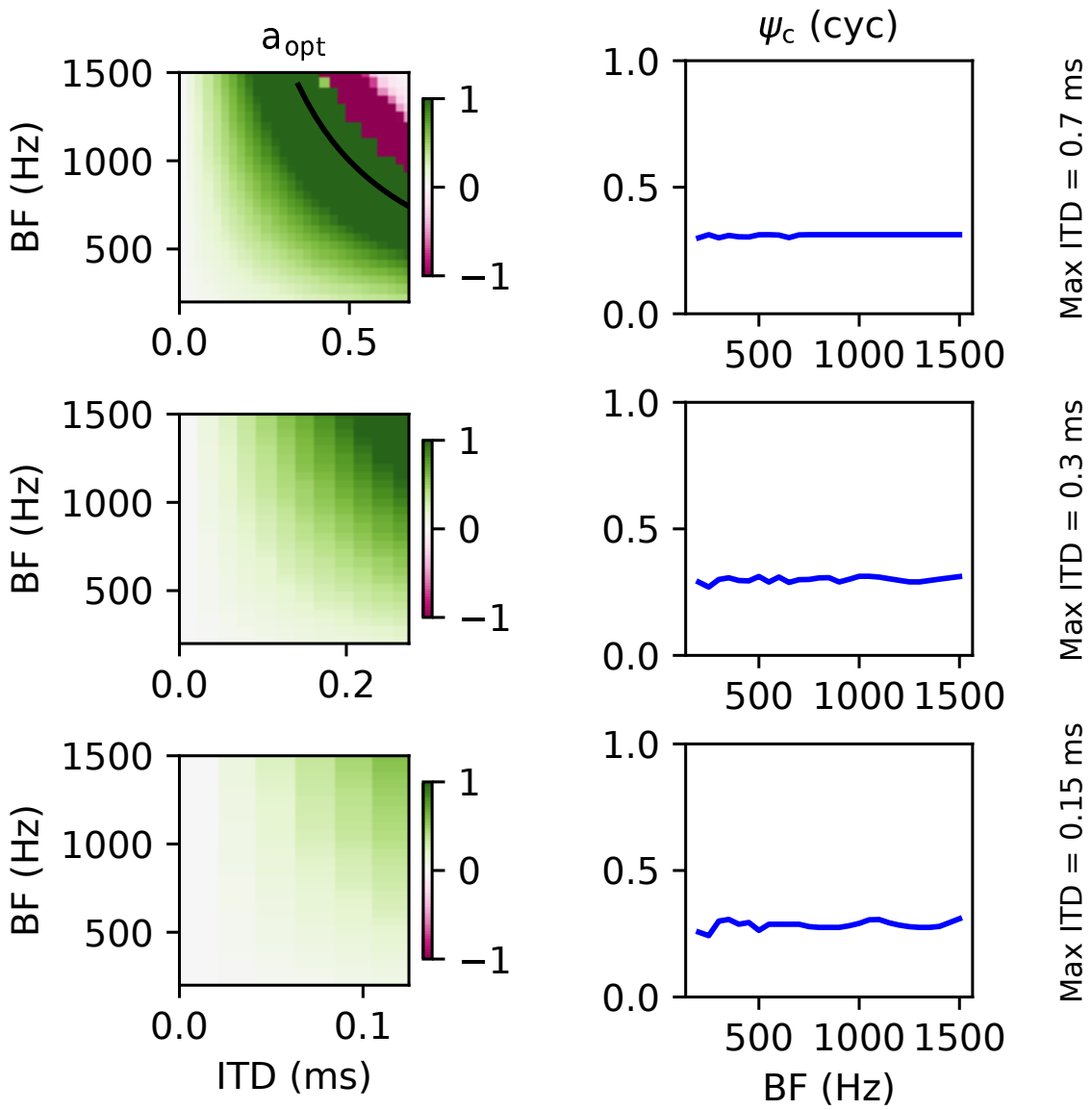

FIG. 3. Optimal parameters. The optimal parameters a (left, color coded; color truncated at $|a| = 1$) and ψ_c (right) for transmitting a stimulus at a given ITD at maximum rate as they were obtained by the optimization procedure described in the main text. Rows show optimal parameters for different maximal physiological ITD (as indicated by x-axis on the left and text on the right).

values of maximal ITDs. The largest ITD bound $\mathrm{ITD}_{\max} = 0.7$ ms roughly corresponds to human head diameters (and will be used for further simulations), 0.3 ms approximates the situation in cats, 0.15 ms roughly matches the physiological ITD range in gerbils. The optimal parameters are remarkably similar for all three choices of $\mathrm{ITD}_{\max}$, with a smooth dependence of the optimal a on ITD and BF. Negative LSO weights $a < 0$ only seem necessary for large head sizes and high BF. Interestingly, the best commissural phase is at

around 0.3 cycles under all conditions.

In order to understand the mechanisms underlying the optimal parameter choice, the input space of the IC (MSO and LSO activity) is examined in Fig. 4 for pure tone stimuli at the BF of three frequency channels and varying ITDs. The trajectories in input space strongly depend on the commissural phase. For the optimal value $\psi_c = 0.3$ cyc (Fig. 4A) and low ITDs (blueish colors) the input trajectories in the MSO-LSO plane have zero angle relative to the MSO axis, whereas high ITDs (redish colors) tend towards steeper positive angles. The difference in angles is more pronounced the higher BF. The angle in the MSO-LSO plane thus encodes stimulus ITD and can be read out by a (normalized) synaptic weight vector

$$\mathbf{w} = \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} 1 \\ a \end{pmatrix} \tag{8}$$

adjusted to a specific ITD (greenish lines in Fig. 4) using the optimal a parameters from Fig. 3. For non-optimal $\psi_c = 0.7$ cyc (Fig. 4B) the ITD separate towards negative angles in the input space. Since, however, the contralateral LSO is excitatory, negative relative weights a cannot be established. Also taking into account the ipsilateral LSO activity (Fig. 4C), does not solve the problem: the ITDs now separated along positive angle in the MSO-LSO space, however, the ipsilateral LSO exerts inhibitory derive to IC neurons and thus the relative weight a cannot be positive. The optimal weight is thus largely $a = 0$, except for very high ITDs and high frequencies, at which strong negative weight evokes the optimal IC response.

Thus, the optimal relative LSO weight a in each frequency channel allows an IC neuron to tune its maximal responsiveness to a specific ITD. The optimal commissural phase ϕ_c determines the shape of the trajectories in MSO-LSO space. For $\psi_c \approx 0.3$ cyc, the trajectories are confined to positive angles that increase with ITD. A further advantage of decoding via an angle in MSO-LSO space is level-invariance: Scaling both signals by the same factor (i.e., making the sound louder or softer) would not change the angle and thus also not require a different decoder.

So far the model is designed entirely feed forward. The IC, however, also hosts recurrent connections [20]. In this paper the recurrent IC network is assumed to provide lateral inhibition within a BF channel. To this end $\mathrm{IC}_\tau(\omega)$ is considered as the input to the local IC

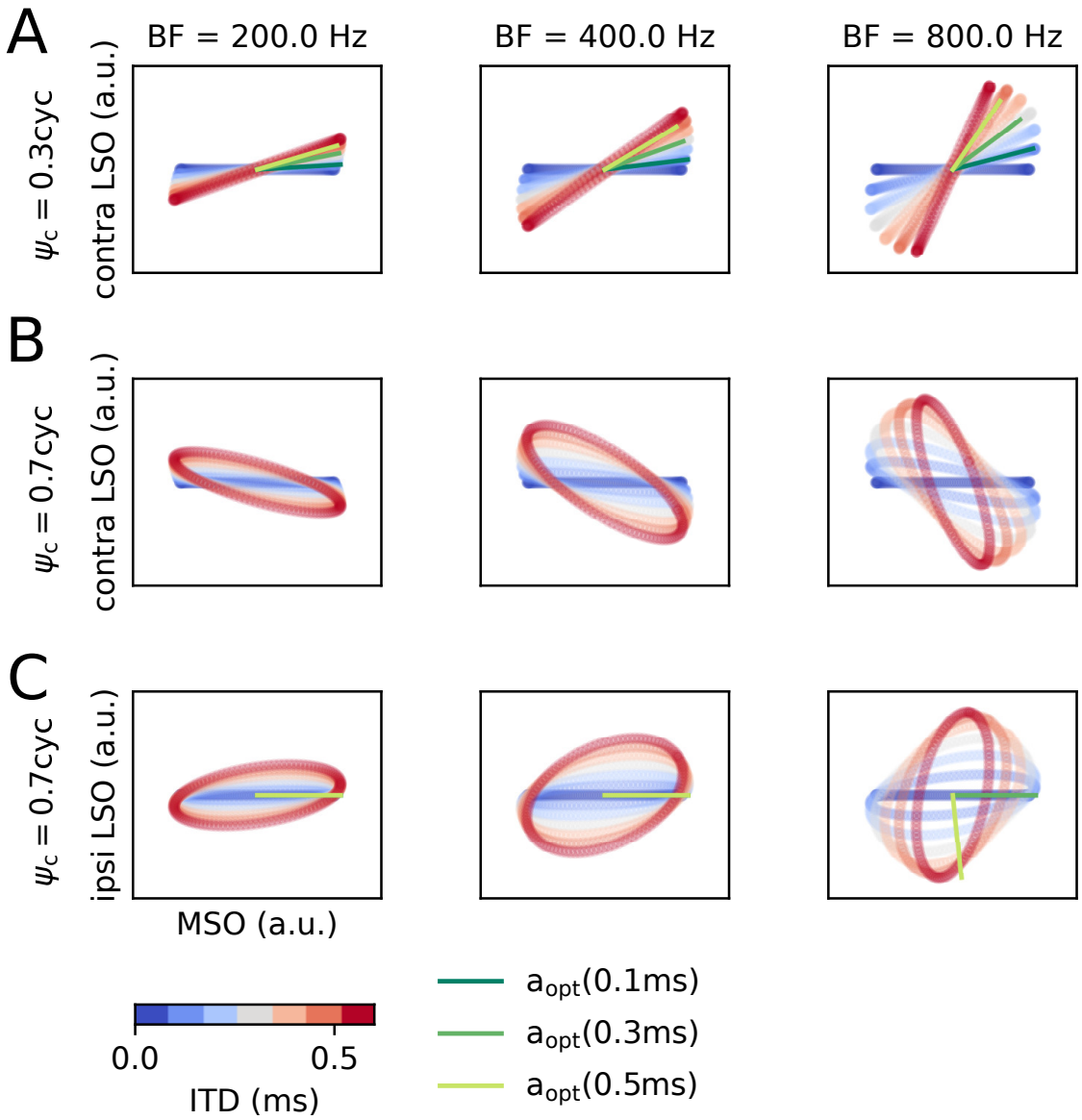

FIG. 4. Brainstem code. The input space (LSO vs. MSO) is plotted for three BF channels (columns as indicated), when stimulated with a pure tone at BF. (A) For a commissural phase $\psi_c = 0.3$ cyc the ITD (blue to red colors) changes the inclination of the trajectories. The optimal weight a_{opt} from Fig. 3 reflects the direction in MSO-LSO space (greenish color) that matches the inclination of the corresponding ITD. (B) For $\psi_c = 0.7$ cyc ITDs are reflected by negative inclinations, which could only be sampled by negative weights a, but are forbidden since the contralateral LSO is excitatory. (C) Using the inhibitory ipsilateral LSO as a second input dimension again results in a positive inclination for $\psi_c = 0.7$ cyc, however, being an inhibitory projection, only negative weights a are allowed. The green directions reflect the optimal (non-positive) a maximizing the response amplitudes.

network, which yields the output activity $\mathrm{IC}_\tau(\omega)^{LI}$ by normalization with the mean input

$$\mathrm{IC}_\tau^{LI}(\omega) = \mathrm{IC}_\tau(\omega)\,[1 + J/N(\omega)\sum_{\tau'}\mathrm{IC}_{\tau'}(\omega)]^{-1}$$

with $N(\omega)$ denoting the number of neurons in the BF channel. All simulations were performed for $J = 20$. The main effect of lateral inhibition is thus to sharpen the activity peak along the ITD (τ) axis according to divisive normalization. In this paper, non-linear compression is only applied to firing rate plots, but not to measures based on sound reconstructions (Pearson's correlations), which require linearity.

SIMULATIONS

First, single cell ITD tuning curves and best ITDs were checked for consistency with experimental data. Although the model was not optimized for generating a broad coverage of ITDs, but rather to maximize firing rate for any given ITD (which still would allow an even higher rate for a different ITD), tuning curve peaks cover a broad range (Fig. 5A). Best ITDs decrease with BF (Fig. 5B) with mean best interaural phase difference at around 0.125 cycles (Fig. 5C) as found experimentally [15]. Hence the distribution of best ITDs matches the empirically found contralateral bias.

Next, the model was tested on its capability to follow dynamically varying ITDs, although the parameters were not optimized for ITD encoding. For comparison with experimental data from (Siveke et al., 2008), "phase warp" stimuli were applied, which are broad band noise stimuli that evoke a strong impression of acoustic motion at the predetermined frequency warp. The simulated population activity indeed reflected the acoustic motion (Figure 5A) and also the broad-band nature of the stimulus (Figure 5B). To account for processing in the low-frequency pathway, stimuli were low pass filtered with a 1st order zero-phase Butterworth filter with edge frequency 0.5 kHz. Phase-locking to the warp frequency, as measured by vector strength,

$$\mathrm{VS} = |\int \mathrm{d}t\, IC(t)^2\, \exp(2\pi \mathrm{i} f_{\mathrm{warp}} t)| / \int \mathrm{d}t\, IC(t)^2\ ,$$

was observed up to 512 Hz, similarly as reported from electrophysiological measurements (Siveke et al., 2008). Also the numerical VS values roughly matched experimental values.

After supporting the model design by comparison with dynamic ITD processing, a classical binaural detection paradigm was applied: Binaural unmasking has been defined as the

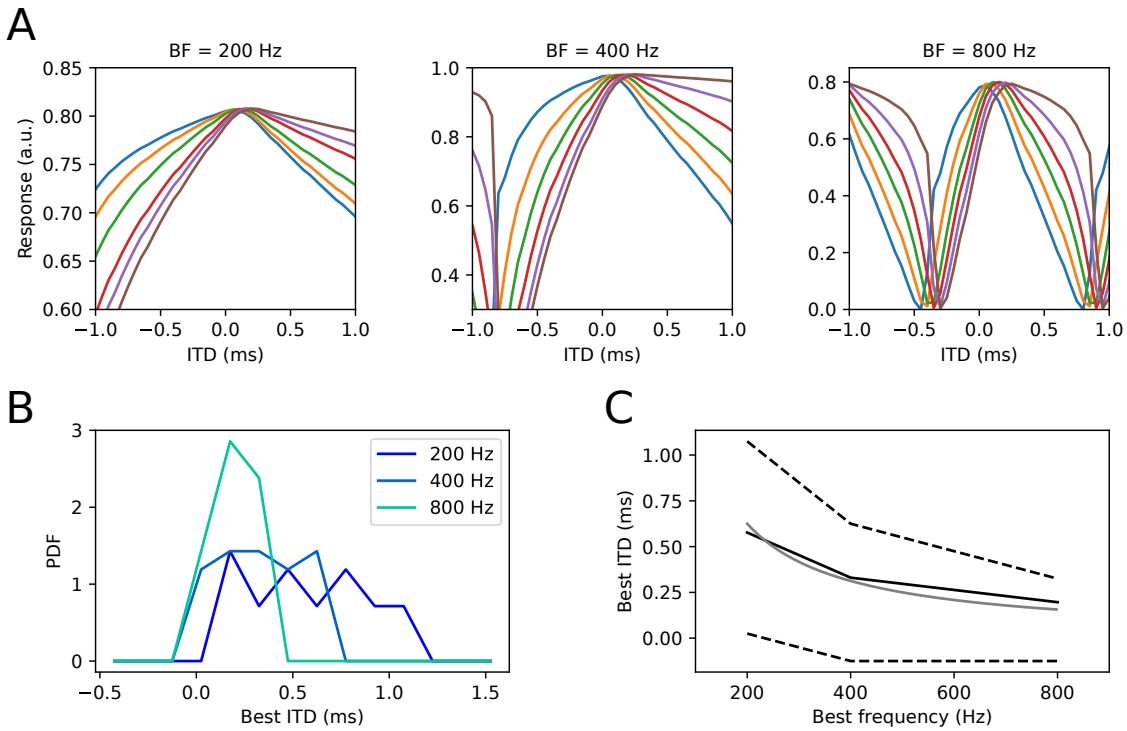

FIG. 5. Best ITDs. (A) Example tuning cures in the 200, 400 and 800 Hz channels obtained with pure tone stimulation at BF. (B) Probability density functions (PDFs) of best ITDs for BF channels as indicated by color. Best ITDs were derived from the phase of the Fourier mode at BF. (C) Mean (solid) and mid-80 percentile (dashed) of best ITDs. Grey line indicates 0.125 cycles. To have results comparable to published data from Guinea pig IC (Agapiou and McAlpine, 2008) parameters were used for a maximum ITD of 0.3 ms; middle panel in Fig. 3.

improvement of signal detection thresholds by presenting signal and masker with different interaural phases (as compared to presenting them with the same phase), without additional amplitude cues (as e.g., ILDs or gap listening) [31]. The difference in detection threshold is called binaural masking level difference (BMLD). To generate interaural phase differences, traditionally either signal ($S_\pi N_0$) or masker ($S_0 N_\pi$) have been sign inverted. In Fig. 7, the model outputs (see Appendix C) are evaluated for the classical $S_0 N_0$ vs. $S_\pi N_0$ situation. All model inputs are low pass-filtered to account for ITD processing being constrained to low-frequency hearing. If both stimulus (400 Hz pure tone) and masker (broadband noise) are presented with 0 phase delay (left column), a reduction of the signal-to-noise ratio (SNR) selectively reduces the Pearson correlation between signal and model output. The output of a Jeffress-like delay line (crosses; see Appendix A) has been simulated for comparison with the theoretically optimal solution.

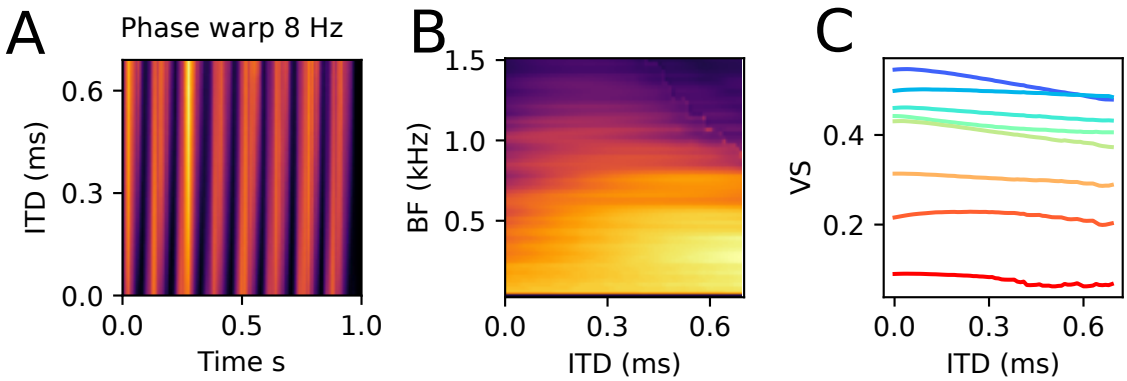

FIG. 6. Fast ITD sensitivity. (A) Population rate as a function of time separated into neuron populations optimized towards the same ITD and averaged over all frequency bands. Color code was normalized between 0 (dark) and maximum (bright). (B) Same simulation as in A averaged over time instead of BF (same 0-max color code). (C) Phase-locking to warp frequency as a function of optimal ITD and warp frequency (color) measured by vector strength (VS).

For an SNR of about -25dB the signal is still clearly visible in the power plots of both S_0N_0 and $S_\pi N_0$, whereas for -30dB, visual inspection no longer allows for clear distinction between signal and noise components in the power plots. Power at the signal frequency, however, shows a small bias towards higher optimal ITDs in the $S_\pi N_0$ condition, suggesting that binaural separation of the signal frequency facilitates detection, as shown in psychoacoustics [30].

While these simulations can provide a qualitative basis for the biological mechanisms underlying the BMLD, quantitative explanation of BMLDs are much harder. Besides the large discrepancy of BMLD values reported in the literature [31], this difficulty also arises because the perceptual thresholds involve additional higher-order processes that "read" the IC population pattern. Despite these caveats, BMLD threshold between -10 and -30 dB are in the range reported in experiments [31].

To evaluate the effect of lateral inhibition, simulations were repeated in a pure feed forward setting ($J = 0$); Fig. 8. While response patterns between S_0N_0 and $S_\pi N_0$ conditions clearly separate in terms of best ITD (x-axis), the detection thresholds for low SNRs (between -25 and -30) seem much less distinct between the two conditions than for simulations with lateral inhibition. Thus the model proposes BMLDs to fundamentally arise from negative feedback between neurons of distinct best ITDs.

So far, the model was only tested for artificial sounds, i.e., pure tones and broad band

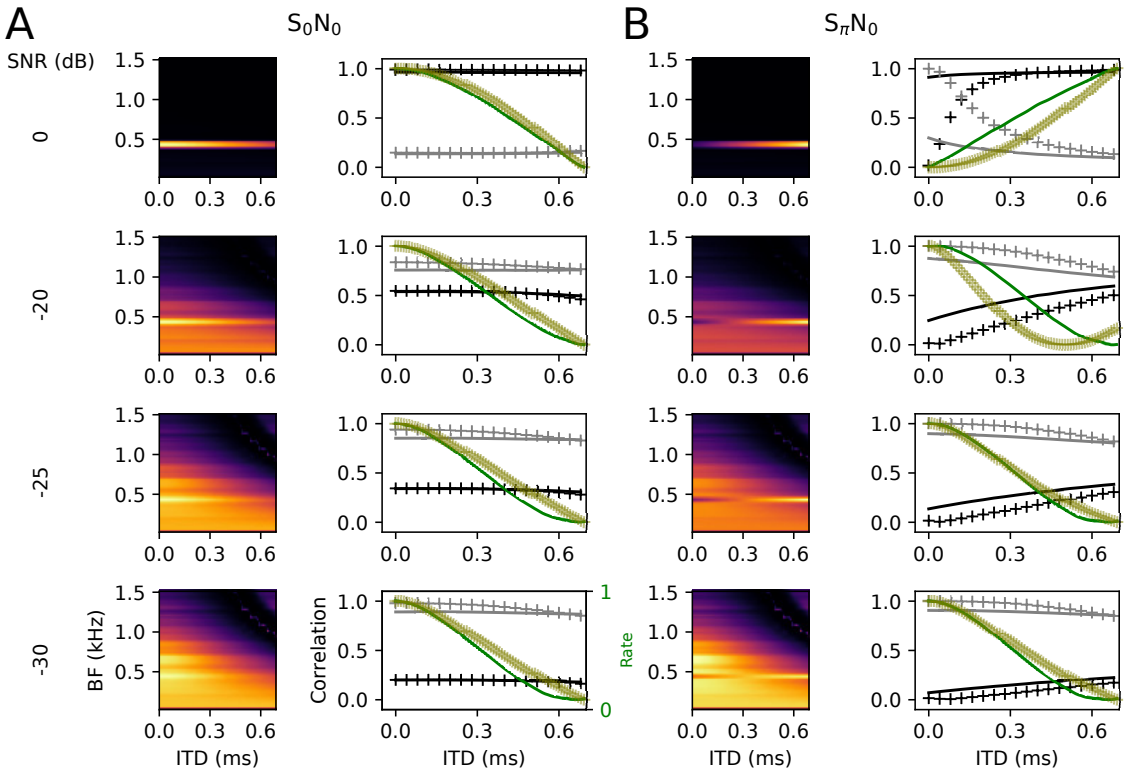

FIG. 7. Binaural unmasking. A 400 Hz pure tone signal (S) and a white noise masker (N) are presented with binaural phase 0 for the noise and binaural phases 0 (A) and π (B) for the tone. The sound level of the signal is presented at different levels (rows), with SNR=0dB indicating that tone and noise have the same broad band intensity. Left: Power $|\mathrm{IC}(BF)|^2$ of the IC model as function of BF channel and ITD for which a was optimized according to Fig. 3. The color code is normalized between 0 (dark) and maximum (bright). Right: The rate of the IC model (green: time average of $\mathrm{IC}(t)$) is plotted as a function of optimal ITD. Rate of the optimal delay line model is shown as olive crosses. The Pearson correlations of the broad band model outputs with the signal and noise are plotted in black and gray, respectively. Solid lines are derived from the IC model, crosses from the optimal delay line model. All results are obtained after low pass filtering the stimuli with a 1st order zero-phase Butterworth filter with edge frequency 0.5 kHz to account for low frequency sound processing. Correlation values are the maximal Pearson's r obtained for all relative shifts between model output and original sound signal in an interval $\pm 2\times$ITD.

noise. Moreover, using a pure tone as signal and noise as masker induces strong spectral asymmetries. Thus, to test the model's applicability to the separation of natural sounds, and to have signal and noise stimuli with comparable spectral content, human speech signals were used as next model inputs (Figure 9). Moreover, the classical $S_{0,\pi}N_{0,\pi}$ stimuli can be

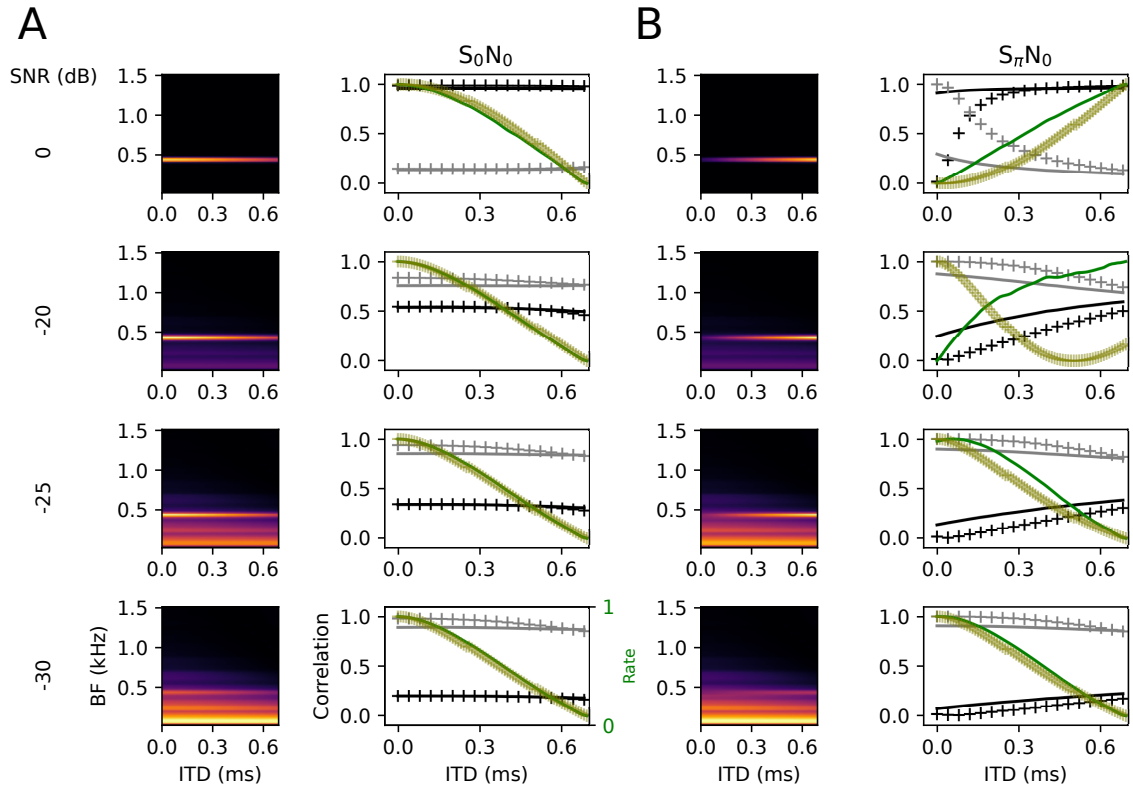

FIG. 8. Binaural unmasking without lateral inhibition. Same as Fig. 7 with lateral coupling set to $J = 0$.

criticized as unphysiological, since an interaural phase difference of π does not correspond to a real sound location for frequencies below approximately 700 Hz. Also for wide band noises maskers, as well as broad band natural sounds, a sign inversion cannot be mapped to a position in space. Binaural unmasking, however, was also shown to exist for "real" ITDs [52] representing a physiologically more realistic sound separation scenario. Therefore, in Fig. 9, sound sources were assigned a sound location via a physiological ITD.

A male speaker was considered as the background (noise) and a female speaker as the signal with varying SNR. For SNR=0, both stimuli were normalized to the same mean overall power. Both signals were placed at physiological ITDs, 0 and 0.4 ms, as indicated in the Figure. The spectral content of both sounds is clearly more similar than in Fig. 7, however, the stimuli are still different, both in spectral content and temporal envelope. The population rate pattern (green) again reflects the ITDs of the sound mixture, although the IC model seems generally less responsive to the signal sound. The correlation signals are strongest at ITDs which are not occupied by the other speaker. A major difference, however, compared to the $S_{0,\pi}N_0$ stimuli is a stark reduction of the threshold level at which

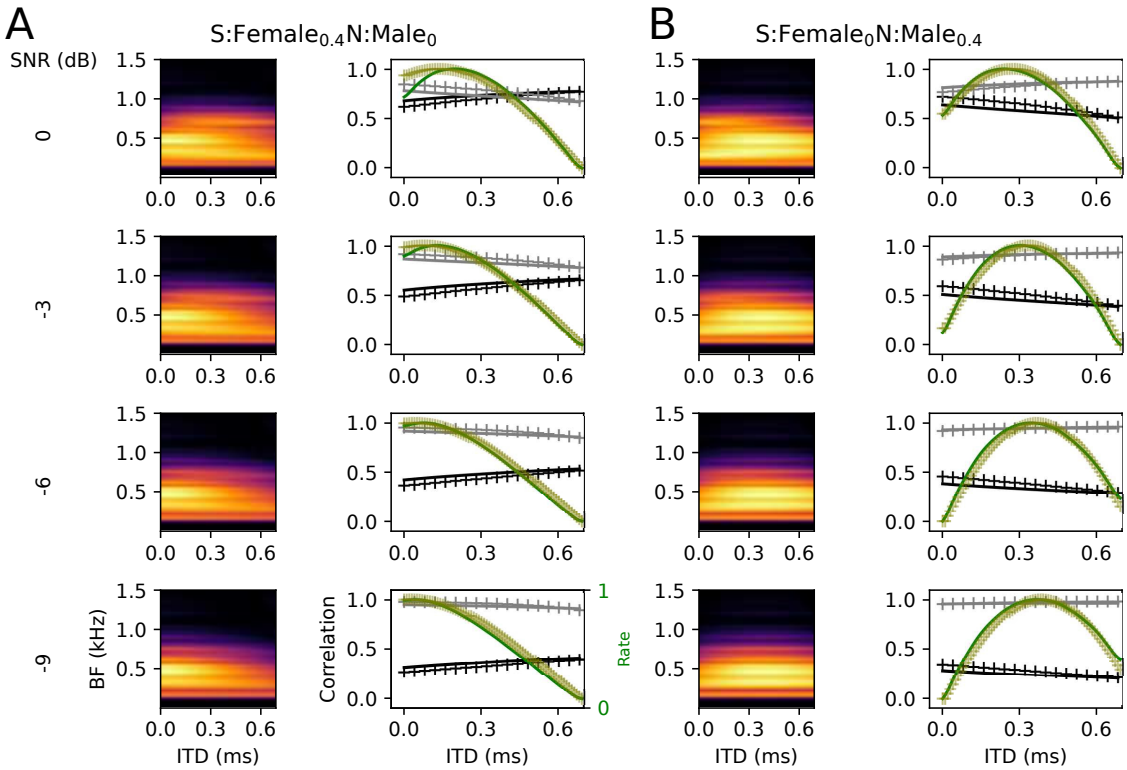

FIG. 9. Binaural unmasking of speech signals. Same as in Fig. 7 with signal and noise replaced by speech sounds from a female and a male speaker, respectively. Note the SNR scale is reduced as compared to Figs. 7 and 8. Speech signals were taken from a free audiobook (https://librivox.org/the-adventures-of-huckleberry-finn-version-5-dramatic-reading-by-mark-twain/). The excerpts are available as supplementary material [53].

the correlations at signal ITD are equal. Also the SNR at which the firing rate peaks at stimulus ITD disappear is strongly reduced. Thus, having comparable spectral content in signal and masker reduces BMLDs to only few dB in both models.

Again, Jeffress-like delay lines and the IC model with optimized parameters a and ψ_c produce almost identical results, showing a near optimal performance of the IC model for speech stimuli.

All results so far have been derived for low-pass filtered stimuli in order to reflect auditory processing in the low-frequency brainstem pathway. The theoretical derivations, however, at no point require a restriction to low frequencies. Particularly the Jeffress-like delay line architecture implements optimal reconstruction independent of the spectral content of the signal. To show that the IC model does not critically rely on the specific cutoff frequency of 2.5 kHz, stimulus separation, as measured by Pearson correlation model output and

signal source, was next tested for mixtures of 2 and 3 speakers without low-pass filtering (Figure 10). Correlation coefficients were only mildly reduced in the IC model as compared to

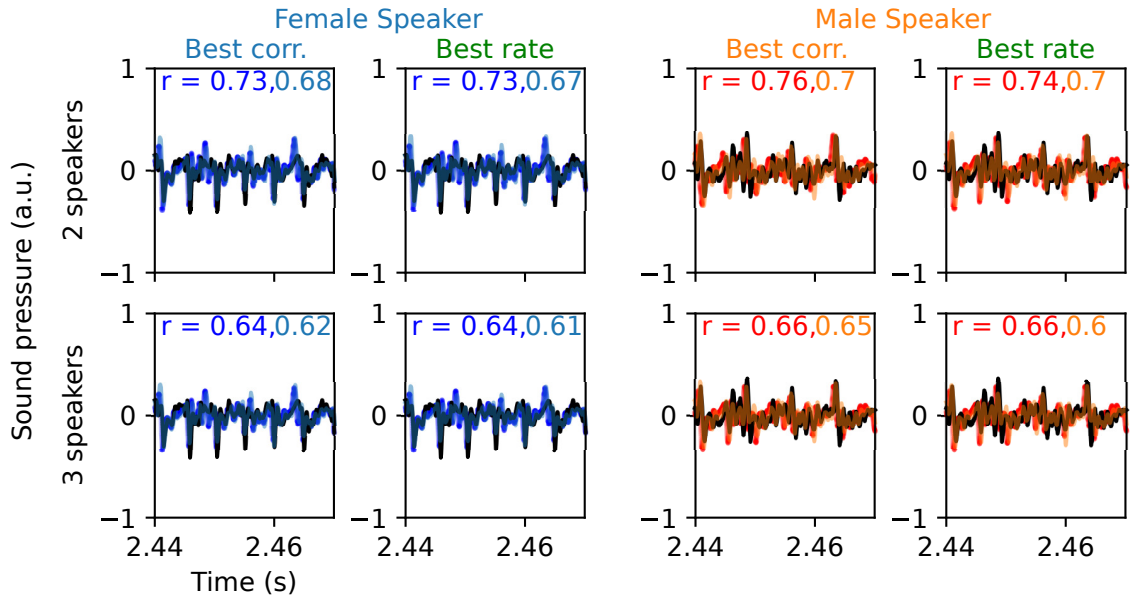

FIG. 10. Deciphering a speech scene. The broad band sound signal (black) was recovered as the model output for the optimal delay line model (dark blue and red) and the IC model (light blue and orange) at the neuron with best ITD generating the highest correlation (Best corr), or the highest rate (Best rate) for a two and a three speaker mixture (Top and bottom row as indicated). Only a small 30 ms snippet is shown. The Pearson's r values are derived from the whole 10 s signals. The three speaker mixture contained another male voice at the same mean intensity. Speech signals were the same as in Figure 9 (Supplementary material [53]).

the optimal reconstruction with delay lines, suggesting that the auditory midbrain circuitry achieves robust almost optimal sound separation performance independent of the specific phase-locking limit and the number of noise sources. For acoustic verification, audio samples of the reconstructed wave forms [54], the mixtures [55], as well as the original sounds [53] can be found as Supplementary Materials.

In a final set of simulations, the model was extended by an adaptive mechanism, assuming that the SOC inputs show firing rate accommodation as reported from measurements [56, 57]. Adaptation was implemented by multiplicative attenuation factors $\alpha_{\mathrm{MSO/LSO}}$ following first order kinetics with steady state 1,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = (1-\alpha_X)/\tau_a - \alpha_X\, c_X\, |X|(t)$$

where X is either MSO or LSO. The adapted SOC inputs are then obtained by

$$X(t) \leftarrow X(t)(b + \alpha_X X)/(b + 1)$$

in which $b/(b+1)$ implements a minimal baseline transmission. Simulations have been obtained for $b = 0.1, \tau_\alpha = 2$s, $c_{MSO} = 5/\tau_\alpha$, and $c_{LSO} = 10/\tau_\alpha$ (with the input signal normalized to maximum amplitude 1). Simulation outcomes are shown in Figure 11 for a pure tone stimulus and a female speaker applied to two contralaterally leading ITDs. Owing to the adaptation mechanism, over time, the angle of the input trajectories in the MSO-LSO plane moves towards the diagonal (Figure 10A), which generates a trend that neurons with intermediate optimal ITDs have largest activity. In sounds with strong amplitude modulations such as speech (Figure 10C) adaptation effects, however, decay during pauses and are no longer present at transients, indicating that ITD information is best preserved at onsets of transients, as reported psychoacoustically [58]. Such adaptation-induced drift in the population code recruits more neurons (i.e., neurons within a broader range of optimal ITDs) and thus could convey sound information with higher population rate during periods with little amplitude modulations at the cost of losing ITD information.

DISCUSSION

A circuit-level model was presented on how the inferior colliculus processes binaural information computed in the superior olivary complex (SOC). For an optimal commissural phase delay ψ_c in the stria of Held of about 0.3 cycles, IC neurons can generate a broad range of best ITDs of the contralateral hemisphere by only adjusting the relative weight of the LSO input. Commissural phases far off 0.3 cycles cannot generate an ITD rate code in the IC given the physiological synaptic kinetics. The present model performs close to optimal in sound source reconstruction, as it closely matches the performance of the theoretically optimal delay-line solution. Generating diverse best ITDs via modulations of synaptic weights is biologically more efficient than using a range of delay lines, because it allows for dynamic adaptations on short time scales. Moreover, systematically varying delay lines have been shown to be absent in the mammalian brainstem [40].

Current neurobiological circuit models of binaural hearing have a strong focus on explaining sound localization, including its mechanisms [50, 59–64], encoding [13, 25, 28, 29, 59, 65] and perception [26, 66]. Sound localization and binaural unmasking, however, involve the

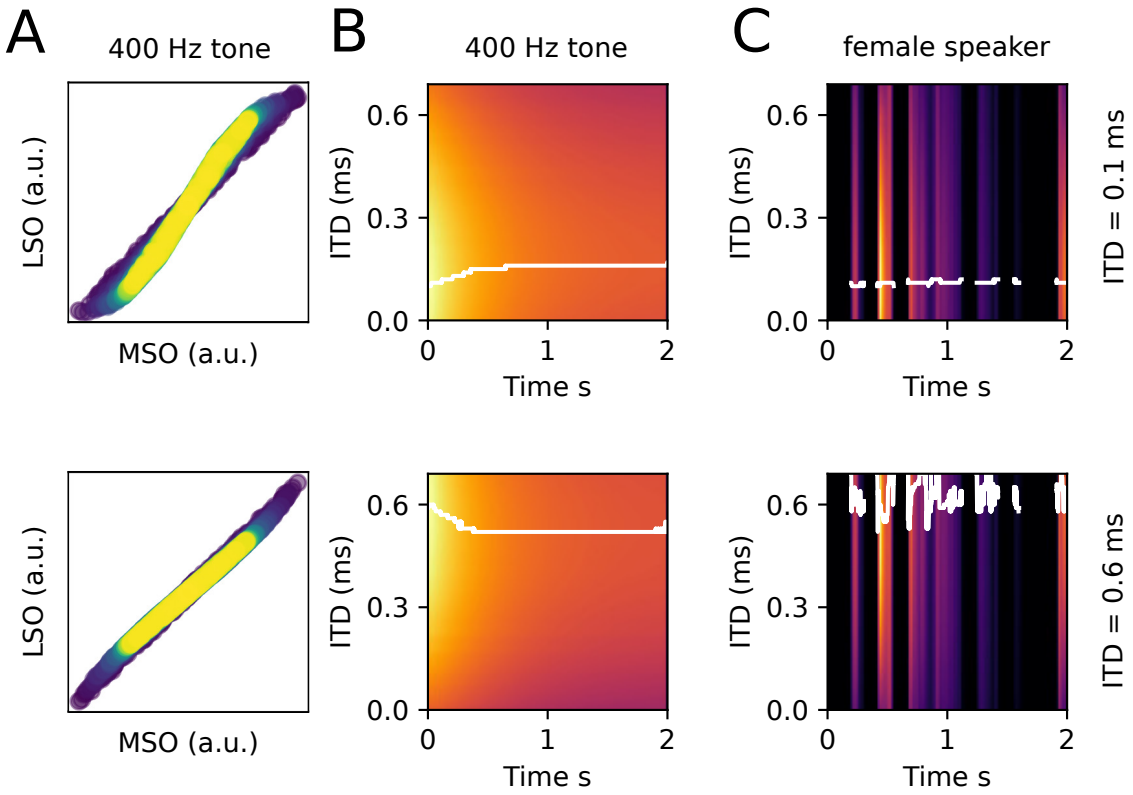

FIG. 11. (A) MSO-LSO input plane (see Figure 3) upon stimulation with a 2 second 400 Hz pure tone with ITD=0.1 ms (top) and ITD=0.6 ms (bottom). The color code reflects the stimulus time (blue early, yellow late). (B) IC population activity (optimal ITD of IC neuron on y-axis) stimulus ITD=0.1 ms (top) and stimulus ITD=0.6 ms (bottom). The color code is normalized between minimum (dark) and maximum (bright). The white transparent line marks the neuron with the maximum activity for any point in time. (C) Same as B for a speech sound from a female speaker.

same neuronal brainstem structures MSO and LSO. Yet, biologically inspired models of binaural IC are scarce. As a notable exception, the work of Dietz et al. (2008) has a strong focus on biological plausibility of the peripheral and SOC processing, however, makes no specific attempt to connect to the IC circuitry. The present model fills this gap in that it largely disregards the biological stages up to the SOC, but emphasizes the contribution of the feed-forward pathways to the IC. This approach necessarily imposes limitations: 1) The effects of peripheral processing are reduced to Fourier decomposition into BF channels and a low pass filtering of the signals to account for the phase-locking limit. Any non-linear effects of peripheral processes are thus neglected, specifically compression and half-wave rectification. Compression, however, could be compensated by level adaptation in the auditory nerve [67] and the brainstem [68] reducing the level dependence of firing rates. Half-wave rectification

may be compensated for by the convergence of a large number of synaptic fibers at the level of the SOC neurons [69]. 2) Detection and localization thresholds cannot be quantitatively predicted by the model. Apart from the errors introduced by the missing peripheral model this is also because of the missing readout model, i.e., how the cortex translates the IC rate code into a percept.

The simplicity of the present model, however, is also its strength, as it has very little unconstrained parameters and as it is derived from the optimization principle of maximizing the response of an IC cell to a sound at a specific ITD. Particularly, avoiding a non-linear periphery model allows for a direct comparison of the reconstructed sound signal (without lateral inhibition) with the original sound wave in the time domain and thus also a direct comparison with the optimal reconstruction model (delay line model), which also is formulated in the time domain.

The present model furthermore resolves a shortcoming of the simple two/four-channel model [24, 26, 66], which does not provide a conceivable explanation for attentional effects. Mammals are able to focus their attention on a specific sound location of behavioral importance [70–72], however, neural mechanisms of attention typically suggest a bias towards neurons with receptive fields in the attended area [73]. Such a bias can be easily mechanistically implemented by facilitating the response or transmission of neurons with "attended" receptive field [74], whereas it is hard to conceive such a bias for a push-pull code of space made out of relative spike counts in the whole neural population. The model thus argues for a labeled line code arising in the IC due to synaptic mechanisms that then can be subject to attention-guided readout. Although the model would work across the whole broad-band spectrum, it is meant to apply only to low frequency brainstem processing, since only there, phase-locked action potentials convey fine structure information of the sound wave form [75, 76]. Interaural level differences (ILDs) are not relevant for binaural unmasking at low stimulus frequencies but contribute sound separation and localization at high frequencies, particularly when ILDs are generalized to broadband short-term envelope fluctuations [31, for review]. While ILDs of high-frequency sounds are used for spatial hearing by all mammals, fine structure ITD information, is virtually absent in auditory nerve activity in mammals with exclusive high-frequency hearing or with very small head sizes. During evolution, however, early air-borne sound hearing mammals were such high-frequency hearers [77]. Also some contemporary bat species have too small heads to use ITDs, but nevertheless possess a monaural MSO [78]. Therefore the general assumption that the evolu-

tion of the MSO was driven by the demand to localize sounds seems not fully convincing and allows for the alternate hypothesis that the mammalian ITD circuit has evolved to perform (low-frequency)-sound source separation, with ITD-based sound location encoding evolving as a by-product. An extension of the present model to high frequency hearing (without fine-structure information) is a necessary next step to support this hypothesis.

`acknowledgements` This work was funded by the Deutsche Forschungsgemeinschaft (DFG) under grant SFB 870 (TP B01). The authors would like to thank Benedikt Grothe for discussions.

---

[1] E. Nachmani, Y. Adi, and L. Wolf, Voice separation with an unknown number of multiple speakers, arXiv **2003.01531** (2020).

[2] M. D. Valero, K. E. Hancock, and M. C. Liberman, The middle ear muscle reflex in the diagnosis of cochlear neuropathy, Hear Res **332**, 29 (2016).

[3] J. J. M. Monaghan, J. A. Garcia-Lazaro, D. McAlpine, and R. Schaette, Hidden Hearing Loss Impacts the Neural Representation of Speech in Background Noise, Curr Biol **30**, 4710 (2020).

[4] J. Liu, J. Stohl, and T. Overath, Hidden hearing loss: Fifteen years at a glance, Hear Res **443**, 108967 (2024).

[5] S. van de Par and A. Kohlrausch, Diotic and dichotic detection using multiplied-noise maskers, J Acoust Soc Am **103**, 2100 (1998).

[6] M. A. Stone and B. C. Moore, Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay, Ear Hear **24**, 175 (2003).

[7] R. Beutelmann and T. Brand, Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners, J Acoust Soc Am **120**, 331 (2006).

[8] S. Roth, F. U. ller, J. Angermeier, W. Hemmert, and S. Zirn, Effect of a processing delay between direct and delayed sound in simulated open fit hearing aids on speech intelligibility in noise, Front Neurosci **17**, 1257720 (2023).

[9] B. Grothe, M. Pecka, and D. McAlpine, Mechanisms of sound localization in mammals, Physiol Rev **90**, 983 (2010).

[10] B. Grothe, C. Leibold, and M. Pecka, The Medial Superior Olivary Nucleus: Meeting the Need for Speed., in *The Oxford Handbook of the Auditory Brainstem* (Oxford University Press, 2019) https://academic.oup.com/book/0/chapter/295230120/chapter-ag-

pdf/51308677/book_34649_section_295230120.ag.pdf.

[11] T. C. T. Yin, P. H. Smith, and P. X. Joris, Neural Mechanisms of Binaural Processing in the Auditory Brainstem, Compr Physiol **9**, 1503 (2019).

[12] J. J. Wenstrup, Z. M. Fuzessery, and G. D. Pollak, Binaural neurons in the mustache bat's inferior colliculus. II. Determinants of spatial responses among 60-kHz EI units, J Neurophysiol **60**, 1384 (1988).

[13] D. McAlpine, D. Jiang, and A. R. Palmer, A neural code for low-frequency sound localization in mammals, Nat Neurosci **4**, 396 (2001).

[14] P. X. Joris, B. Van de Sande, D. H. Louage, and M. van der Heijden, Binaural and cochlear disparities, Proc Natl Acad Sci U S A **103**, 12917 (2006).

[15] J. P. Agapiou and D. McAlpine, Low-frequency envelope sensitivity produces asymmetric binaural tuning curves, J Neurophysiol **100**, 2381 (2008).

[16] D. McAlpine, D. Jiang, T. M. Shackleton, and A. R. Palmer, Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus, J Neurosci **18**, 6026 (1998).

[17] D. C. Fitzpatrick, S. Kuwada, and R. Batra, Transformations in processing interaural time differences between the superior olivary complex and inferior colliculus: beyond the Jeffress model, Hear Res **168**, 79 (2002).

[18] A. Klug, E. E. Bauer, J. T. Hanson, L. Hurley, J. Meitzen, and G. D. Pollak, Response selectivity for species-specific calls in the inferior colliculus of Mexican free-tailed bats is generated by inhibition, J Neurophysiol **88**, 1941 (2002).

[19] D. W. Grantham and F. L. Wightman, Detectability of varying interaural temporal differences, J Acoust Soc Am **63**, 511 (1978).

[20] L. Yassin, M. Pecka, J. Kajopoulos, H. Gleiss, L. Li, C. Leibold, and F. Felmy, Differences in synaptic and intrinsic properties result in topographic heterogeneity of temporal processing of neurons within the inferior colliculus, Hear Res **341**, 79 (2016).

[21] P. X. Joris, B. van de Sande, A. Recio-Spinoso, and M. van der Heijden, Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation, J Neurosci **26**, 279 (2006).

[22] I. Siveke, S. D. Ewert, B. Grothe, and L. Wiegrebe, Psychophysical and physiological evidence for fast binaural processing, J Neurosci **28**, 2043 (2008).

[23] T. M. Shackleton and A. R. Palmer, The time course of binaural masking in the inferior colliculus of guinea pig does not account for binaural sluggishness, J Neurophysiol **104**, 189

(2010).

[24] N. A. Lesica, A. Lingner, and B. Grothe, Population coding of interaural time differences in gerbils and barn owls, J Neurosci **30**, 11696 (2010).

[25] H. Lüling, I. Siveke, B. Grothe, and C. Leibold, Frequency-invariant representation of interaural time differences in mammals, PLoS Comput Biol **7**, e1002013 (2011).

[26] A. Lingner, M. Pecka, C. Leibold, and B. Grothe, A novel concept for dynamic adjustment of auditory space, Sci Rep **8**, 8335 (2018).

[27] S. Groß, *Sound source coding in the azimuthal plane: separating sounds via short-term interaural time difference estimations*, Ph.D. thesis, Ludwig-Maximilians-Universität München (2021).

[28] N. S. Harper and D. McAlpine, Optimal neural population coding of an auditory spatial cue, Nature **430**, 682 (2004).

[29] D. F. Goodman, V. Benichoux, and R. Brette, Decoding neural responses to temporal cues for sound localization, Elife **2**, e01312 (2013).

[30] I. J. Hirsh, The influence of interaural phase on interaural summation and inhibition, The Journal of the Acoustical Society of America **20**, 536 (1948).

[31] J. F. Culling and M. Lavandier, Binaural unmasking and spatial release from masking, in *Binaural Hearing: With 93 Illustrations*, edited by R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper (Springer International Publishing, Cham, 2021) pp. 209–241.

[32] M. van der Heijden and P. X. Joris, Interaural correlation fails to account for detection in a classic binaural task: dynamic ITDs dominate N0Spi detection, J Assoc Res Otolaryngol **11**, 113 (2010).

[33] M. Dietz, S. D. Ewert, V. Hohmann, and B. Kollmeier, Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences, Brain Research **1220**, 234 (2008), active Listening.

[34] M. Dietz and G. Ashida, Computational models of binaural processing, in *Binaural Hearing: With 93 Illustrations*, edited by R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper (Springer International Publishing, Cham, 2021) pp. 281–315.

[35] L. R. Bernstein and C. Trahiotis, The normalized correlation: Accounting for binaural detection across center frequency, The Journal of the Acoustical Society of America **100**, 3774 (1996), https://pubs.aip.org/asa/jasa/article-pdf/100/6/3774/12163033/3774_1_online.pdf.

[36] S. van de Par and A. Kohlrausch, A new approach to comparing binaural masking level

differences at low and high frequencies, J Acoust Soc Am **101**, 1671 (1997).

[37] R. Wan, N. I. Durlach, and H. S. Colburn, Application of a short-time version of the Equalization-Cancellation model to speech intelligibility experiments with speech maskers, J Acoust Soc Am **136**, 768 (2014), https://pubs.aip.org/asa/jasa/article-pdf/136/2/768/14869645/768_1_online.pdf.

[38] T. Biberger and S. D. Ewert, Towards a simplified and generalized monaural and binaural auditory model for psychoacoustics and speech intelligibility, Acta Acust. **6**, 23 (2022).

[39] R. M. Stern and C. Trahiotis, Models of binaural interaction, Handbook of perception and cognition **6**, 347 (1995).

[40] S. Karino, P. H. Smith, T. C. Yin, and P. X. Joris, Axonal branching patterns as sources of delay in the mammalian auditory brainstem: a re-examination, J Neurosci **31**, 3016 (2011).

[41] C. Kapfer, A. H. Seidl, H. Schweizer, and B. Grothe, Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons, Nat. Neurosci. **5**, 247 (2002).

[42] K. Couchman, B. Grothe, and F. Felmy, Medial superior olivary neurons receive surprisingly few excitatory and inhibitory inputs with balanced strength and short-term dynamics, J Neurosci. **30**, 17111 (2010).

[43] M. T. Roberts, S. C. Seeman, and N. L. Golding, A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry, Neuron **78**, 923 (2013).

[44] T. J. Park, B. Grothe, G. D. Pollak, G. Schuller, and U. Koch, Neural delays shape selectivity to interaural intensity differences in the lateral superior olive, J Neurosci **16**, 6554 (1996).

[45] D. J. Tollin and T. C. Yin, Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive, J Neurosci **25**, 10648 (2005).

[46] B. Beiderbeck, M. H. Myoga, N. I. C. Müller, A. R. Callan, E. Friauf, B. Grothe, and M. Pecka, Precisely timed inhibition facilitates action potential firing for spatial coding in the auditory brainstem, Nat Commun **9**, 1771 (2018).

[47] T. P. Franken, B. J. Bondy, D. B. Haimes, J. H. Goldwyn, N. L. Golding, P. H. Smith, and P. X. Joris, Glycinergic axonal inhibition subserves acute spatial sensitivity to sudden increases in sound intensity, eLife **10**, e62183 (2021).

[48] K. K. Glendenning, B. N. Baker, K. A. Hutson, and R. B. Masterton, Acoustic chiasm V: inhibition and excitation in the ipsilateral and contralateral projections of LSO, J Comp Neurol **319**, 100 (1992).

[49] S. Brückner and R. Rübsamen, Binaural response characteristics in isofrequency sheets of the

gerbil inferior colliculus, Hear Res **86**, 1 (1995).

[50] A. Brand, O. Behrend, T. Marquardt, D. McAlpine, and B. Grothe, Precise inhibition is essential for microsecond interaural time difference coding, Nature **417**, 543 (2002).

[51] B. Grothe and T. J. Park, Time can be traded for intensity in the lower auditory system, Naturwissenschaften **82**, 521 (1995).

[52] T. L. Langford and L. A. Jeffress, Effect of Noise Crosscorrelation on Binaural Signal Detection, The Journal of the Acoustical Society of America **36**, 1455 (1964), https://pubs.aip.org/asa/jasa/article-pdf/36/8/1455/11601549/1455_1_online.pdf.

[53] See supplemental material at [url will be inserted by publisher] for original sounds. sounds are excerpts from https://librivox.org/the-adventures-of-huckleberry-finn-version-5-dramatic-reading-by-mark-twain/.

[54] See supplemental material at [url will be inserted by publisher] for reconstructed sounds.

[55] See supplemental material at [url will be inserted by publisher] for sound mixtures.

[56] A. K. Magnusson, T. J. Park, M. Pecka, B. Grothe, and U. Koch, Retrograde GABA signaling adjusts sound localization by balancing excitation and inhibition in the brainstem, Neuron **59**, 125 (2008).

[57] A. Stange, M. H. Myoga, A. Lingner, M. C. Ford, O. Alexandrova, F. Felmy, M. Pecka, I. Siveke, and B. Grothe, Adaptation in sound localization: from GABA(B) receptor-mediated synaptic modulation to perception, Nat. Neurosci. **16**, 1840 (2013).

[58] M. Dietz, T. Marquardt, N. H. Salminen, and D. McAlpine, Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds, Proc Natl Acad Sci U S A **110**, 15151 (2013).

[59] L. A. Jeffress, A place theory of sound localization, J Comp Physiol Psychol **41**, 35 (1948).

[60] C. Leibold and J. L. van Hemmen, Spiking neurons learning phase delays: how mammals may develop auditory time-difference sensitivity, Phys Rev Lett **94**, 168102 (2005).

[61] Y. Zhou, L. H. Carney, and H. S. Colburn, A model for interaural time difference sensitivity in the medial superior olive: interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology, J Neurosci **25**, 3046 (2005).

[62] P. E. Jercog, G. Svirskis, V. C. Kotak, D. H. Sanes, and J. Rinzel, Asymmetric excitatory synaptic dynamics underlie interaural time difference processing in the auditory system, PLoS Biol **8**, e1000406 (2010).

[63] C. Leibold, Influence of inhibitory synaptic kinetics on the interaural time difference sensitivity

in a linear model of binaural coincidence detection, J Acoust Soc Am **127**, 931 (2010).

[64] S. Lehnert, M. C. Ford, O. Alexandrova, F. Hellmundt, F. Felmy, B. Grothe, and C. Leibold, Action potential generation in an anatomically constrained model of medial superior olive axons, J Neurosci **34**, 5370 (2014).

[65] G. von Bekesy, Zur Theorie des Hörens; über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schallwirkungen, Physik Zeitschr **31**, 824–835, 857–868 (1930).

[66] T. M. Shackleton, B. C. Skottun, R. H. Arnott, and A. R. Palmer, Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of Guinea pigs, J Neurosci **23**, 716 (2003).

[67] B. Wen, G. I. Wang, I. Dean, and B. Delgutte, Dynamic range adaptation to sound level statistics in the auditory nerve, J Neurosci **29**, 13797 (2009).

[68] B. D. B. Willmore and A. J. King, Adaptation in auditory processing, Physiol Rev **103**, 1025 (2023).

[69] A. R. Callan, M. Heß, F. Felmy, and C. Leibold, Arrangement of Excitatory Synaptic Inputs on Dendrites of the Medial Superior Olive, J Neurosci **41**, 269 (2021).

[70] G. Rhodes, Auditory attention and the representation of spatial information, Percept Psychophys **42**, 1 (1987).

[71] C. C. Lee and J. C. Middlebrooks, Auditory cortex spatial sensitivity sharpens during task performance, Nat Neurosci **14**, 108 (2011).

[72] M. H. Myoga, D. Amaro, V. Kello, M. Gumbert, C. Leibold, M. Pecka, and B. Grothe, The dynamics of context-dependent space representations in auditory cortex, bioRxiv 10.1101/2023.11.25.568638 (2023), https://www.biorxiv.org/content/early/2023/11/27/2023.11.25.568638.full.pdf.

[73] A. Borji and L. Itti, Optimal attentional modulation of a neural population, Front Comput Neurosci **8**, 34 (2014).

[74] J. H. Reynolds and D. J. Heeger, The normalization model of attention, Neuron **61**, 168 (2009).

[75] P. H. Smith, P. X. Joris, L. H. Carney, and T. C. Yin, Projections of physiologically characterized globular bushy cell axons from the cochlear nucleus of the cat, J Comp Neurol **304**, 387 (1991).

[76] P. X. Joris and T. C. Yin, Responses to amplitude-modulated tones in the auditory nerve of

the cat, J Acoust Soc Am **91**, 215 (1992).

[77] B. Grothe and M. Pecka, The natural history of sound localization in mammals – a story of neuronal inhibition, Front Neural Circuits **8**, 116 (2014).

[78] B. Grothe, E. Covey, and J. H. Casseday, Medial superior olive of the big brown bat: Neuronal responses to pure tones, amplitude modulations, and pulse trains, Journal of Neurophysiology **86**, 2219 (2001), https://doi.org/10.1152/jn.2001.86.5.2219.

[79] R. Penrose, A generalized inverse for matrices, Mathematical Proceedings of the Cambridge Philosophical Society **51**, 406–413 (1955).

Appendix A: Optimal Signal Reconstruction

To find the optimal estimate of a sound signal in background noise the mixture matrix A is transformed to the frequency domain

$$\begin{pmatrix} \hat{R}(\omega) \\ \hat{L}(\omega) \end{pmatrix} = A \begin{pmatrix} \hat{S}(\omega) \\ \hat{N}_R(\omega) \\ \hat{N}_L(\omega) \end{pmatrix}$$

with

$$A = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\omega\Delta t/2} & 1 & 0 \\ \mathrm{e}^{-\mathrm{i}\omega\Delta t/2} & 0 & 1 \end{pmatrix} .$$

The mean square optimal reconstruction [27] is obtained by the Moore-Penrose pseudo inverse $A^* = \lim_{\delta\to 0}(A^\dagger A + \delta \mathbb{1})^{-1} A^\dagger$ [79], with $\dagger$ denoting the Hermitian conjugate. The frequency representation in the limit $\delta \to 0$,

$$A^* = \frac{1}{3} \begin{pmatrix} \mathrm{e}^{-\mathrm{i}\omega\Delta t/2} & \mathrm{e}^{\mathrm{i}\omega\Delta t/2} \\ 2 & -\mathrm{e}^{\mathrm{i}\omega\Delta t} \\ -\mathrm{e}^{-\mathrm{i}\omega\Delta t} & 2 \end{pmatrix} ,$$

corresponds to the time domain representation of the optimal estimates

$$\tilde{S}(t) = \frac{1}{3}\left[R\left(t - \frac{\Delta t}{2}\right) + L\left(t + \frac{\Delta t}{2}\right)\right] \tag{A1}$$

$$\tilde{N}_R(t) = \frac{1}{3}\left[2\,R(t) - L(t + \Delta t)\right] \tag{A2}$$

$$\tilde{N}_L(t) = \frac{1}{3}\left[2\,L(t) - R(t - \Delta t)\right] \tag{A3}$$

Appendix B: IC model

The IC model from Eq. (6)

$$IC(t) = [R(t) - a\,R(t-\Lambda_l-\Lambda_c) + L(t-\Lambda_m) + a\,L(t-\Lambda_c)]/\sqrt{1+a^2}$$

is transformed to the frequency domain as

$$\widehat{IC}(\omega) = [\hat{R}(\omega)\,(1 - a\,\mathrm{e}^{-\mathrm{i}(\psi_l+\psi_c)}) + \hat{L}(\omega)\,(\mathrm{e}^{-\mathrm{i}\psi_m} + a\,\mathrm{e}^{-\mathrm{i}\psi_c})]/\sqrt{1+a^2} \tag{B1}$$

With $\hat{R} = \hat{S}\,\mathrm{e}^{-\mathrm{i}\Delta\phi/2}$ and $\hat{L} = \hat{S}\,\mathrm{e}^{\mathrm{i}\Delta\phi/2}$ the amplitude of the IC response in frequency band ω equals

$$\begin{aligned}
\widehat{IC}(\omega) &= \hat{S}(\omega)\left[\mathrm{e}^{\mathrm{i}\Delta\phi/2}(1 - a\,\mathrm{e}^{-\mathrm{i}(\psi_l+\psi_c)}) + \mathrm{e}^{-\mathrm{i}\Delta\phi/2}(\mathrm{e}^{-\mathrm{i}\psi_m} + a\,\mathrm{e}^{-\mathrm{i}\psi_c})\right]/\sqrt{1+a^2} \\
&= \hat{S}(\omega)\left[\mathrm{e}^{-\mathrm{i}\Delta\phi/2}\,\gamma_R + \mathrm{e}^{\mathrm{i}\Delta\phi/2}\,\gamma_L\right]/\sqrt{1+a^2} \\
&= \hat{S}(\omega)\,\mathrm{e}^{\mathrm{i}\Delta\phi/2}\,\mathrm{e}^{\mathrm{i}\arg\gamma_R}\,\left[|\gamma_R| + |\gamma_L|\mathrm{e}^{-\mathrm{i}(\Delta\phi+\Delta\Phi)}\right]
\end{aligned} \tag{B2}$$

with

$$\gamma_R(a) = 1 - a\,\exp[-\mathrm{i}(\psi_l+\psi_c)] \tag{B3}$$

$$\gamma_L(a) = \exp(-\mathrm{i}\Delta\phi)\,[\exp(-\mathrm{i}\psi_m) + a\,\exp(-\mathrm{i}\psi_c)] \tag{B4}$$

$$\tag{B5}$$

and the phase difference $\Delta\Phi = \arg[\gamma_R/\gamma_L] = \arg[\gamma_R\,\bar{\gamma_L}]$, as announced in the Results section.

For the ipsilateral LSO, the argument can be made analogously with

$$\gamma_R^{(i)}(a) = 1 + a \tag{B6}$$

$$\gamma_L^{(i)}(a) = \exp(-\mathrm{i}\Delta\phi)[\exp(-\mathrm{i}\psi_m) - a\,\exp(-\mathrm{i}\psi_l)]\ . \tag{B7}$$

Appendix C: Simulations

The IC model output were computed in Fourier space with frequency dependent parameters a and ψ_c from Fig. 3. In detail, left and right ear sounds were sampled with 96 kHz. Broad band sound stimuli were low-pass filtered by a first order Butterworth filter with cutoff frequency 500Hz unless otherwise mentioned. A fast Fourier transform was applied to snippets of 4001 samples enveloped with a Blackman window. Subsequent snippets were offset by 2000 samples, such that the overlap of snippets was approximately 50%. The fast

Fourier transformed sounds for the left and right ear were then mapped to an IC response according to Eq. (B2) in each frequency channel. After backtransformation into the time domain the windowed snippets were added up with appropriate time shifts corresponding to the 2000 sample points.

All Python code used to produce the results and the figures presented in this paper can be found under www.github.com/cleibold/ICmodel.