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

Christian Leibold[1,2*], Sebastian Groß[3],

1 Fakultät für Biologie & Bernstein Center Freiburg, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
2 BrainLinks-BrainTools Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
3 Fakultät für Biologie, Ludwig-Maximilians Universität München, Munich, Germany

*christian.leiboldbiologie.uni-freiburg.de

Abstract

The inferior colliculus (IC) is a central hub in the ascending auditory brainstem. It hosts many neurons tuned to interaural time differences (ITDs). ITD tuning, however, is already observed and generated one synapse upstream in the superior olivary complex and the physiological mechanisms as well as the functional purpose of the IC projection remain partially unresolved. Here, we argue that combining ITD sensitive inputs from medial superior olive (MSO) and lateral superior olive (LSO) requires a temporally well adjusted delay of cross-hemispheric 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 cross-hemispheric 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 of all sound sources in a complex auditory scene. ITD tuning of IC neurons would then results as a side effect. The model focuses on the low-frequency range, is consistent with the distribution of best ITDs observed in experimental recordings and performs close to optimal in sound source reconstruction.

Author Summary

Natural acoustic scenes consist of many spatially distributed and concurrently active sound sources. We show how such concurrent sounds can be optimally reconstructed solely using information about interaural arrival time differences. We argue that this optimal solution is almost matched by a neural circuit model inspired by the ascending auditory pathway. The model is computationally very efficient because changing only one synaptic weight moves the azimuthal focus of a midbrain neuron. The model predicts a universal phase delay of cross-hemispheric delay lines in the ascending auditory brainstem.

Introduction

Separation of sound sources in noisy acoustic environments is one of the most intriguing capabilities of the human auditory system [1] and still poses a hard problem to modern artificial sound separation solutions [2]. In search for the underlying neural mechanisms

of sound separation and/or detection, binaural hearing turned out to be of fundamental importance [3–5]. 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 [6–8]: 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. [9–12]). 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 [13–15], and even suggest mechanisms of de-novo generation of ILD sensitivity [16], a normative functional approach of binaural representations in the IC based on circuit mechanisms is missing. A particular short-coming thereby arises from the sluggishness assumption [17] 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 [18], 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 [19–21].

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 independent populations, each one of which only conveys the summed activity of all its neurons in a frequency band further downstream [22–28]. 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, 29, 30], 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 [31, 32], and for which ITDs are the dominant cues at low frequencies [33].

Binaural unmasking has so far attracted only little attention of biologically inspired modelling [28, 34, 35], despite highly evolved psychophysical models [36–40]; for review of classical models see [41]. 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.

Models and Methods

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 (an approximation for low-frequency sounds). 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.

Minimizing the mean square error between a linear estimate $\tilde{S}$ of S from the two "measurements" R and L requires inverting the mixture matrix A, which combines equations (1) and (2), and is easiest expressed in the frequency domain, where delays $\Delta t/2$ translate to phases $\mathrm{e}^{\mathrm{i}\omega\Delta t/2}$,

$$\begin{pmatrix} \hat{R}(\omega) \\ \hat{L}(\omega) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathrm{e}^{\mathrm{i}\omega\Delta t/2} & 1 & 0 \\ \mathrm{e}^{-\mathrm{i}\omega\Delta t/2} & 0 & 1 \end{pmatrix}}_{A} \begin{pmatrix} \hat{S}(\omega) \\ \hat{N}_R(\omega) \\ \hat{N}_L(\omega) \end{pmatrix} \tag{3}$$

The mean square optimal reconstruction [27] is obtained by the Moore-Penrose pseudo inverse $A^* = \lim_{\delta\to 0}(A^\dagger\, A + \delta\mathbb{1})^{-1}\, A^\dagger$ [42], with $\dagger$ denoting the Hermitian conjugate. The frequency representation in the limit $\delta \to 0$,

$$A^* = \frac{1}{3} \begin{pmatrix} \mathrm{e}^{-\mathrm{i}\omega\Delta t/2} & \mathrm{e}^{\mathrm{i}\omega\Delta t/2} \\ 2 & -\mathrm{e}^{\mathrm{i}\omega\Delta t} \\ -\mathrm{e}^{-\mathrm{i}\omega\Delta t} & 2 \end{pmatrix} , \tag{4}$$

corresponds to the time domain representation of the optimal estimates

$$\tilde{S}(t) = \frac{1}{3}\left[R\left(t - \frac{\Delta t}{2}\right) + L\left(t + \frac{\Delta t}{2}\right)\right] \tag{5}$$

$$\tilde{N}_R(t) = \frac{1}{3}\left[2\,R(t) - L(t + \Delta t)\right] \tag{6}$$

$$\tilde{N}_L(t) = \frac{1}{3}\left[2\,L(t) - R(t - \Delta t)\right] \tag{7}$$

Eq. 5 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.

Delay lines, however, have not been found in the mammalian sound localization circuitry [43], and thus alternative solutions are of interest of how Eq. (5) can be implemented in ways more consistent with mammalian brainstem anatomy.

Sensitivity to ITDs and submillisecond synaptic integration has not only been reported for MSO neurons, but also for LSO neurons [47–50]. 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

$$MSO(t) = R(t) + L(t - \Lambda_m) \tag{8}$$

$$LSO(t) = L(t) - R(t - \Lambda_l) , \tag{9}$$

with $\Lambda_{m/l}$ denoting the cross-hemispheric latencies. Notably, R and L, strictly speaking are no longer pressure waves but neural activity. For the sake of analytical tractability we, however, assume that phase-locked neural inputs approximately match the sound wave if summed over a large set of axonal fibers, thereby interpreting negative values as below spontaneous activity. High numbers of synaptic release sites per axon at a single MSO cell justify this approximation [51] .

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,

$$IC(t) = [MSO(t) + a\,LSO(t - \Lambda_c)]/\sqrt{1 + a^2} \tag{10}$$

$$= [R(t) - a\,R(t - \Lambda_l - \Lambda_c) + L(t - \Lambda_m) + a\,L(t - \Lambda_c)]/\sqrt{1 + a^2} , \tag{11}$$

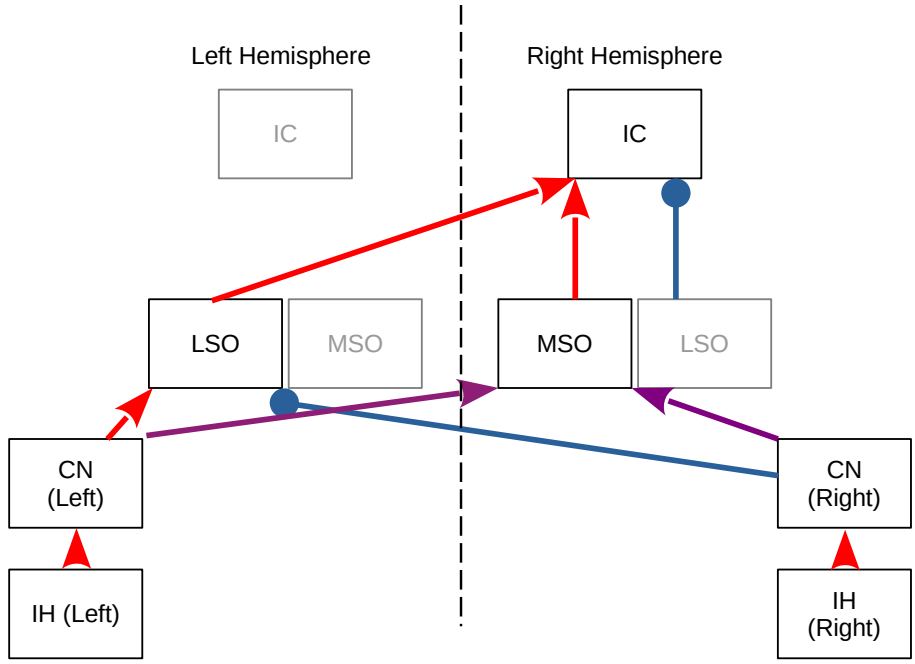

Fig 1. Ascending binaural brainstem anatomy. The inner hair cells (IH) convert mechanical vibrations of the basilar membrane into an electrical signal that is conveyed to the ipsilateral cochlear nuclei (CN) via glutamatergic synaptic transmission (red) of spiral ganglion neurons. CN bushy cell axon 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 cross-hemispheric (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. Notably, the present model design allows for binaural processing already in one hemisphere (see also [26]), which is in contrast to standard two channel approaches as e.g. [22, 28, 54], and which would account for the conserved localization ability in the unaffected hemisphere of patients with midbrain stroke [55, 56].

So far our theory neglects frequency-specific processing, and thus, in order to increase biological realism, Eqs. (8-11) should be considered in a best frequency (BF)-channel dependent manner. The Fourier transform of Eq. (11) reads

$$\widehat{IC}(\omega) = [\hat{R}(\omega)\,(1 - a\,\mathrm{e}^{-\mathrm{i}(\psi_l+\psi_c)}) + \hat{L}(\omega)\,(\mathrm{e}^{-\mathrm{i}\psi_m} + a\,\mathrm{e}^{-\mathrm{i}\psi_c})]/\sqrt{1+a^2}\ . \qquad (12)$$

We next intend to optimize the response of the IC neurons in frequency band ω to an isolated sound source and therefore set $N_{R/L} = 0$. Accordingly, we set $\hat{R} = \hat{S}\,\mathrm{e}^{\mathrm{i}\omega\,\Delta t/2}$

and $\hat{L} = \hat{S}\,\mathrm{e}^{-\mathrm{i}\omega\,\Delta t/2}$, and thus the amplitude of the IC response equals

$$\begin{aligned}
\widehat{IC}(\omega) &= \hat{S}(\omega)\left[\mathrm{e}^{\mathrm{i}\omega\,\Delta t/2}(1 - a\,\mathrm{e}^{-\mathrm{i}(\psi_l+\psi_c)}) + \mathrm{e}^{-\mathrm{i}\omega\,\Delta t/2}(\mathrm{e}^{-\mathrm{i}\psi_m} + a\,\mathrm{e}^{-\mathrm{i}\psi_c})\right]/\sqrt{1+a^2} && (13)\\
&= \hat{S}(\omega)\left[\mathrm{e}^{-\mathrm{i}\omega\Delta t/2}\,\gamma_R + \mathrm{e}^{\mathrm{i}\omega\Delta t/2}\,\gamma_L\right]/\sqrt{1+a^2} && (14)\\
&= \hat{S}(\omega)\,\mathrm{e}^{\mathrm{i}\omega\Delta t/2}\,\mathrm{e}^{\mathrm{i}\arg\gamma_R}\,\left[|\gamma_R| + |\gamma_L|\mathrm{e}^{-\mathrm{i}(\omega\Delta t+\arg(\gamma_R\,\bar{\gamma_L}))}\right] && (15)
\end{aligned}$$

with

$$\begin{aligned}
\gamma_R(a) &= 1 - a\,\exp[-\mathrm{i}(\psi_l+\psi_c)] && (16)\\
\gamma_L(a) &= \exp(-\mathrm{i}\omega\Delta t)\,[\exp(-\mathrm{i}\psi_m) + a\,\exp(-\mathrm{i}\psi_c)]\ . && (17)\\
& && (18)
\end{aligned}$$

The ITD tuning curve of an IC neuron in the frequency channel $\omega = 2\pi$ BF is then modeled by the power of the Fourier coefficient,

$$|\widehat{IC}(\omega)|^2 = |\hat{S}(\omega)|^2\,\left[|\gamma_R|^2 + |\gamma_L|^2 + 2\,|\gamma_R\gamma_L|\,\cos\left(\arg(\gamma_R\,\bar{\gamma_L})\right)\right]/(1+a^2)\ , \quad (19)$$

An 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) [57]; $\psi_l = 0$ [48, 58]) 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 [52], which would change the above formulas to $\gamma_R^{(i)} = 1 + a$, $\gamma_L^{(i)} = \exp(-\mathrm{i}\omega\Delta t)[\exp(-\mathrm{i}\psi_m) - a\exp(-\mathrm{i}\psi_l)]$, and evidently is independent of the commissural phase delay ψ_c.

Importantly, processing in the low-frequency pathway, and particularly the limits imposed by the synaptic integration time constant at IC neurons of about 2 ms [18], is accounted for by low-pass filtering all sound stimuli with a 1st order zero-phase Butterworth filter with edge frequency 0.5 kHz, unless mentioned otherwise.

The dependence of the ITD tuning curves $|IC|^2$ (Eq. 19) 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. Notably, for $a = -1$, the MSO and LSO contribution driven by the right ear input exactly cancel out ($\gamma_R^{(i)} = 1 + a$), and so no ITD information remains and the tuning curve is flat.

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 target ITD and ψ_c, the optimal weight

$$a(\mathrm{ITD}, \psi_c) = \mathrm{argmax}_{a\in[-10,10]}|IC|^2 \quad (20)$$

was chosen to maximize the firing rate of an IC cell according to Eq. (19) for stimuli at a target ITD value. This objective deviates from the usual approach to represent auditory space using a wide range of best ITDs, but rather tries to transmit at maximal

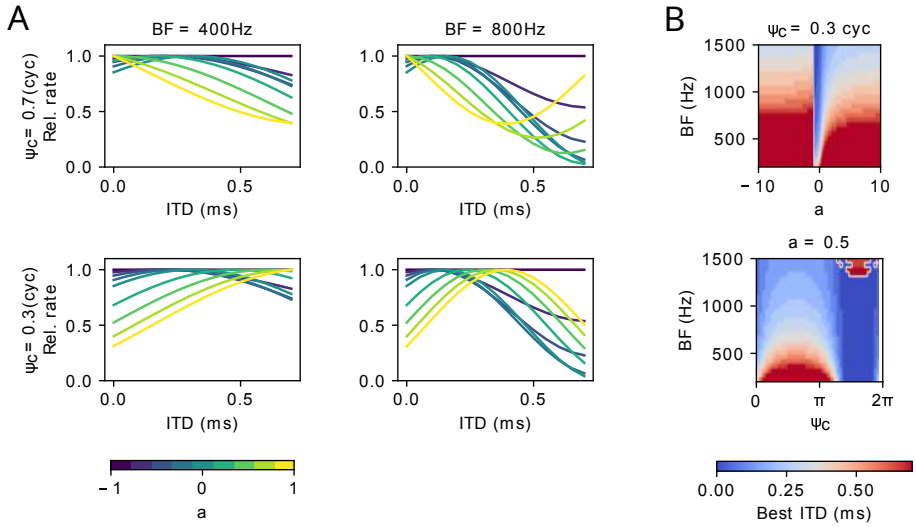

Fig 2. Example ITD tuning curves. (A) ITD tuning curves derived from Eq. (19) are shown for four exemplary conditions and varying relative LSO weights a (colors). Top panels are obtained with a large commissural phase $\psi_c = 0.7$ cyc, bottom panels for small commissural phase $\psi_c = 0.3$ cyc. Left 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.

rate for a signal at the target ITD. In a second step ψ_c, was determined by minimizing the loss function

$$\mathcal{L}(\psi_c) = \int_0^{\mathrm{ITD}_{\max}} \mathrm{dITD}\ \left(|a(\mathrm{ITD}, \psi_c)| - \ln|a(\mathrm{ITD}, \psi_c)| + \lambda\,\mathrm{sign}[-a(\mathrm{ITD}, \psi_c)]\right) \ . \tag{21}$$

By construction, the loss function tries to minimize the mean relative LSO weight $a(\mathrm{ITD})$ (accounting for the relatively small low-frequency LSO), while the $-\ln$ term avoids pure MSO solutions with $a = 0$. The regularization $\lambda\,\mathrm{sign}[-a(\mathrm{ITD}, \psi_c)]$ punishes negative weights a by means of the sign function, i.e., it punishes solutions including the inhibitory ipsilateral pathway, which, from anatomy, is known to be much smaller in size [52]. 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 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 most 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 target 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

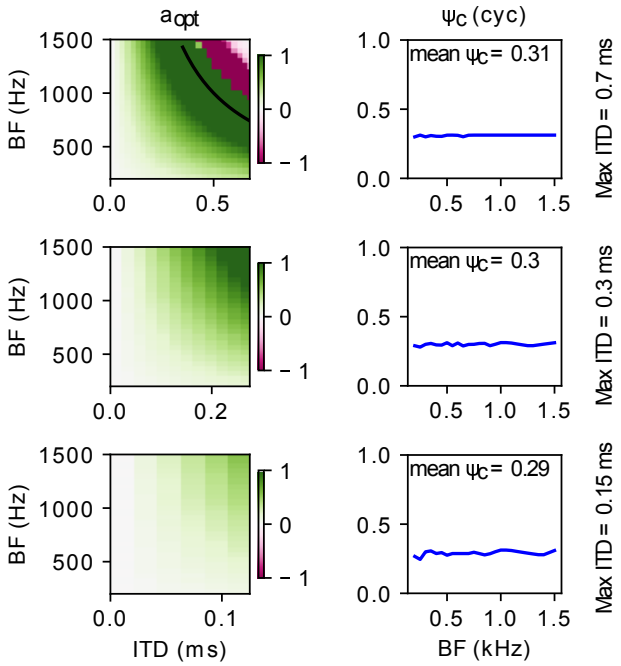

Fig 3. Optimal parameters. The optimal parameters a (left, color coded; color truncated at $|a| = 1$) and ψ_c (right) for transmitting a stimulus at a given ITD at maximum rate as they were obtained by the optimization procedure described in the main text. Rows show optimal parameters for different maximal physiological ITD (as indicated by x-axis on the left and text on the right). The black line in the top left panel indicates the π-limit, i.e., BF$\times$ITD$= 0.5$.

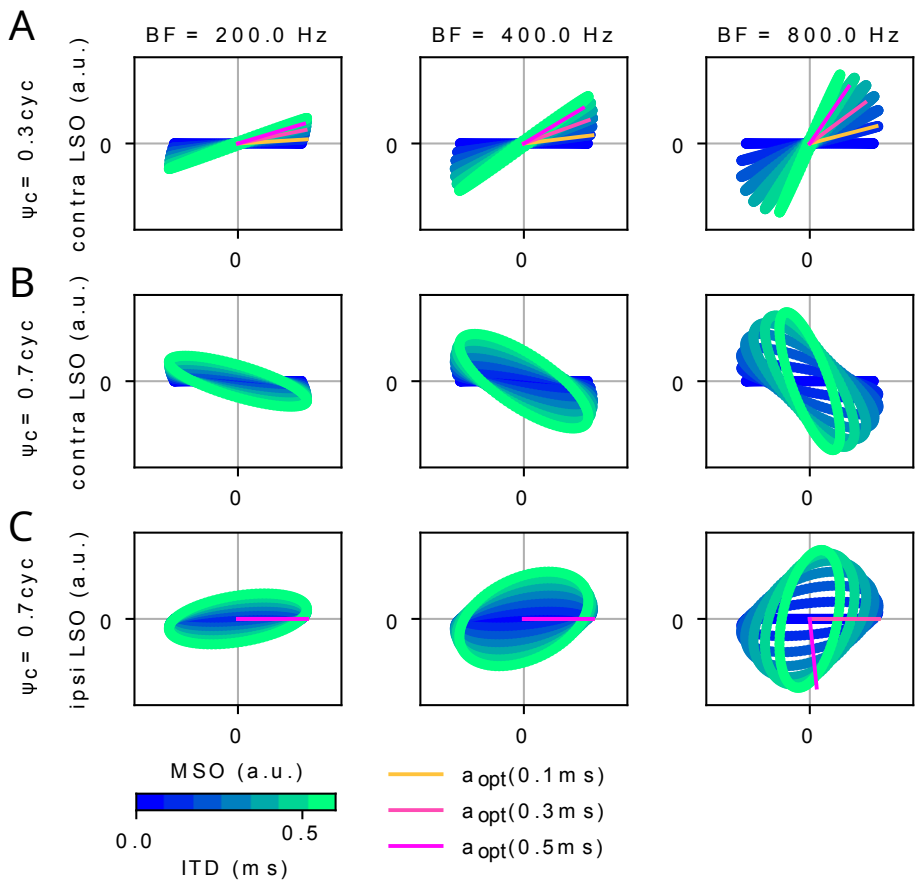

Fig 4. Brainstem code. The input space (LSO vs. MSO) is plotted for three BF channels (columns as indicated), when stimulated with a pure tone at BF. (A) For a commissural phase $\psi_c = 0.3$ cyc the ITD (blue to red colors) changes the inclination of the trajectories. The optimal weight a_{opt} from 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.

ITD and can be read out by a (normalized) synaptic weight vector

$$\mathbf{w} = \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} 1 \\ a \end{pmatrix} \tag{22}$$

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 ψ_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 [18]. In this paper the recurrent IC network is assumed to provide inhibition within a BF channel such that the activity of all neurons is suppressed according to the total amount of activity in the BF channel. To this end, $\mathrm{IC}_\tau(\omega)$ is considered as the input to the IC neuron at atraet ITD τ, which yields the output activity $\mathrm{IC}^N_\tau(\omega)$ by normalization with the mean input

$$\mathrm{IC}^N_\tau(\omega) = \mathrm{IC}_\tau(\omega)\,[1 + \frac{J}{N(\omega)} \sum_{\tau'} \mathrm{IC}_{\tau'}(\omega)]^{-1} \tag{23}$$

with $N(\omega)$ denoting the number of neurons with BF=$\omega/(2\pi)$. All simulations were performed for an inhibitory synaptic weight of $J = 20$. The main effect of inhibition is thus to sharpen the activity peak along the target ITD (τ) axis. In this paper, inhibition is only included to firing rate computations, but not to measures based on sound reconstructions (Pearson's correlations), which require linearity.

For the final set of simulations, we also included rate adaptation, implemented by attenuation variables $\alpha_{\mathrm{MSO/LSO}} < 0$ that multiply the corresponding IC inputs $MSO(t)$ and $LSO(t)$ from eqs. 8. The attenuation variables follow first order kinetics with steady state 1, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\alpha_X = (1-\alpha_X)/\tau_a - \alpha_X\, c_X\, |X|(t) \tag{24}$$

where $X \in \{\mathrm{MSO,LSO}\}$, and $X(t)$ refers to $MSO(t)$ or $LSO(t)$, respectively. The adapted SOC activities are then obtained by

$$X(t) \leftarrow X(t)(b + \alpha_X\, X)/(b+1) \tag{25}$$

in which $b/(b+1)$ implements a minimal baseline activity level for the maximally adapted state $\alpha_X = 0$. 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 details

The IC model outputs 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. (15) 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.

The delay-line model was simulated according to eq. (5).

Stimuli

All sound stimuli used in this study were sampled at 96 kHz. We used a) pure ton stimuli of 200, 300, 400, 600 and 800 Hz, b) broad band white noise, c) speech signals of two male and one female speaker reading passages from free audiobook. The speech excerpts are available as Supporting Information S1.

Results

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 the full physiological range (here 0.3ms; 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 [12]. Hence, although IC neurons integrate LSO and MSO input with additional commissural delay, the distribution of best ITDs matches the empirically found contralateral bias. Expectedly, the shapes of the ITD tuning curves (particularly for lower BFs) deviates from physiological data (e.g., [59]), since the model output $|IC|^2$ rather corresponds to membrane potential amplitudes and the spike rates would be a non-linear transformation of $|IC|^2$. The Best ITD, however, would be unaffected by such a monotonous non-linearity.

Next, the model was tested on its capability to follow dynamically varying ITDs [34], although the parameters were not optimized for ITD encoding. For comparison with experimental data from [20], "phase warp" stimuli were applied, which are a broad band version of a binaural beat that evoke a strong impression of acoustic motion for low warp frequencies f_{warp} and binaural flutter for $f_{\text{warp}} \gtrapprox 5$ Hz. The stimuli were created in frequency space such that for one ear, each frequency (f) component was assigned equal magnitude and a random phase ϕ_f, and for the other ear that phase ϕ_f was assigned to the frequency component shifted upward by a difference f_{warp}. The simulated IC population activity indeed reflected the fastly fluctuating ITDs (Fig. 6A) and also the broad-band nature of the stimulus (Fig. 6B), as described psychophysically and physiologically [20, 33, 34]. Phase-locking to the warp frequency, as measured by vector strength,

$$\text{VS} = |\int \mathrm{d}t\, IC(t)^2\, \exp(2\pi \mathrm{i} f_{\text{warp}} t)| / \int \mathrm{d}t\, IC(t)^2\ , \tag{26}$$

was observed up to 512 Hz, similarly as reported from electrophysiological measurements [20], reflecting the 500 Hz low-pass filtering that is supposed to model

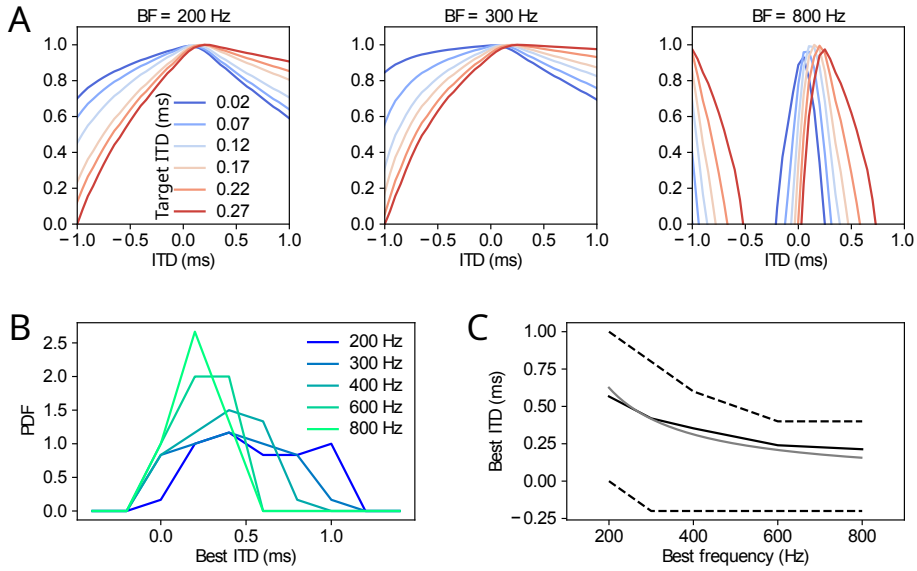

Fig 5. Best ITDs. (A) Example tuning curves in the 200, 400 and 800 Hz channels obtained with pure tone stimulation at BF. Different colors indicate responses of neurons with different Target ITDs (in ms). (B) Probability density functions (PDFs) of *Best ITDs* for BF channels as indicated by color. 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. Tuning curves were obtained for a physiological ITD range of 0.3.ms; middle panel in Fig. 3.

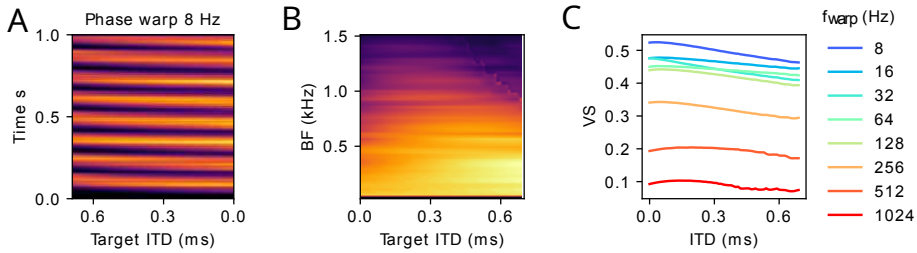

Fig 6. Fast ITD sensitivity. (A) Population rate as a function of time separated into neuron populations optimized towards the same ITD and averaged over all frequency bands. Color code was normalized between 0 (dark) and maximum (bright). (B) Same simulation as in A averaged over time instead of BF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

the postsynaptic current kinetics. Also the numerical VS values roughly matched experimental values of [20] in their Figure 5.

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 improvement of signal detection thresholds by presenting signal and masker with different interaural phases (as compared to presenting them with the same phase at both ears), without additional amplitude cues (as e.g., ILDs or gap listening) [32]. 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 are evaluated for the classical $S_0 N_0$ vs. $S_\pi N_0$ situation. A reduction of the signal-to-noise ratio (SNR)

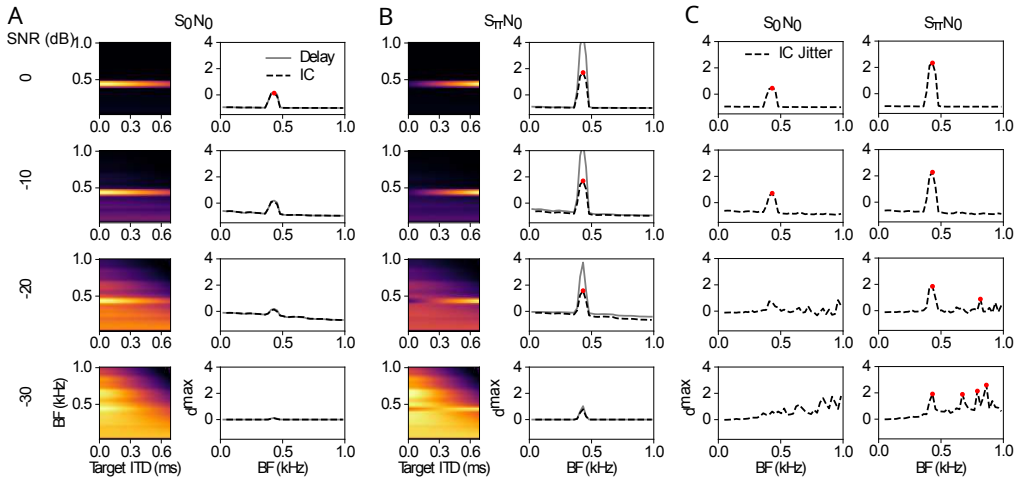

Fig 7. Binaural unmasking. A 400 Hz pure tone signal (S) and a white noise masker (N) are presented with binaural phase 0 for the noise and binaural phases 0 (A) and π (B) for the tone. The sound level of the signal is presented at different levels (rows), with SNR=0dB indicating that tone and noise have the same broad band intensity. Left: Power $|\mathrm{IC}(BF)|^2$ of the IC model as function of BF channel and target ITD for which a was optimized to according to Fig. 3. The color code is normalized between 0 (dark) and maximum (bright). Right: Maximal normalized difference $d^{\max}$ between signal a noise and noise alone in each BF channel. The results for the delay line model are depicted in grey. The red dot indicates the the peak of IC model exceeding a prominence threshold of 1. (C) $d^{\max}$ plots for the two stimulus situations with a Gaussian jitter ($\sigma = 0.2$) applied to the optimal parameters a and ψ_c.

selectively reduces the detectability of the tone signal at 400 Hz. For $\mathrm{S_0N_0}$ the relative signal to noise differences $d^{\max}$ are more sensitive to SNR than for the $\mathrm{S_\pi N_0}$ condition , suggesting that binaural separation of the signal facilitates detection, as shown in psychoacoustics [31].

While these simulations can provide a qualitative basis for the biological mechanisms underlying the BMLD, quantitative explanation of BMLDs are much harder, also because the perceptual thresholds involve additional higher-order processes that "read" the IC population pattern. BMLD thresholds of about -15 to -20 dB are reported in psychoacoustic experiments [60], while BMLDs derived from single neurons are only few dB on average [61, 62]. Thus extracting psychoacoustic threshold requires additional statistical assumptions beyond the neural tuning properties. An estimate for detection thresholds that could be applied to our simulations is the peak prominence of the relative differences $d^{\max}(BF) = \max_{\mathrm{ITD}} \frac{|\mathrm{IC}(BF,ITD)|}{|\mathrm{IC_{noise}}(BF,ITD)|}^2 - 1$ between model output from "tone plus noise" and model output "noise alone" as a function of BF. Peak prominence is defined as the peak amplitude minus the highest local minimum. Setting a prominence threshold of 1 approximately explains experimental BMLDs (red dots in Fig. 7) without attempting to fit all psychoacoustics data. For the delay line model (grey $d^{\max}$ curves) peak prominences in the $\mathrm{S_\pi N_0}$ condition are generally higher, but the threshold of 1 would be crossed at comparable SNR values, yielding similar BMLDs. Notably, we are not proposing peak prominence as a psychophysical readout model, but rather a biologically plausible way of how a population code could be read by downstream neurons (potentially involving multiple stations).

In order to test the sensitivity of the BMLD effect on our optimization procedure (Fig. 3), we repeated the simulation with a 20% random Gaussian jitter around the

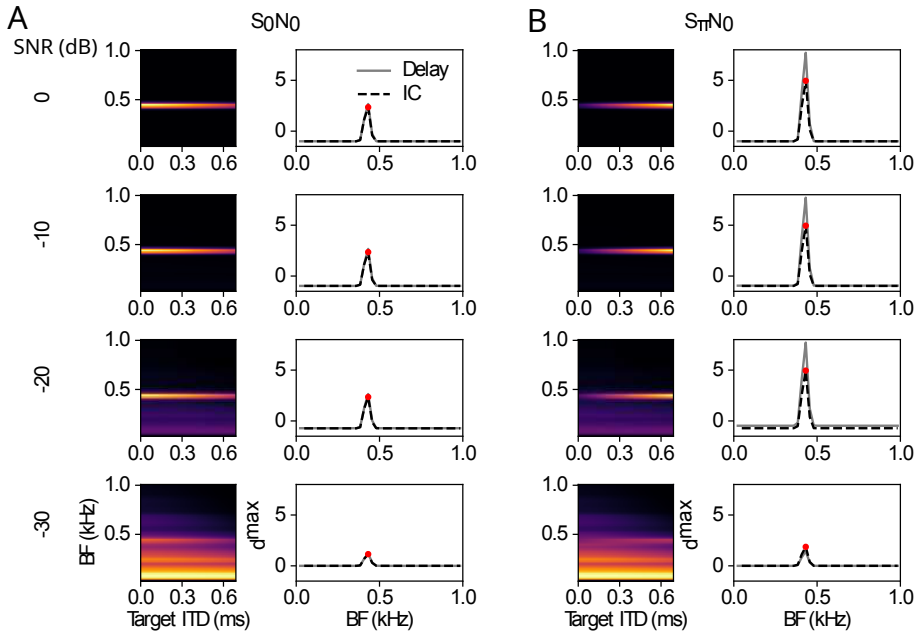

Fig 8. Binaural unmasking without inhibition. Same as Fig. 7 with inhibitory coupling set to $J = 0$.

optimal a and ψ_c values (Fig. 7C). Parameter jitter introduces spurious peaks of $d^{\max}$ and generally reduces detection thresholds beyond psychoacoustic values [60]. Hence, deviations from the optimal parameters have marked effect on detection thresholds.

To furthermore evaluate the effect of 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 target 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 inhibition. Thus the model proposes BMLDs to fundamentally arise from negative feedback between neurons of distinct target ITDs.

The classical $S_{0,\pi}N_{0,\pi}$ stimuli can be 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 noise 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 ITDs [63] representing a physiologically more realistic sound separation scenario. Therefore, in Fig. 9, sound sources were assigned a sound location via a physiological ITD, 0 and 0.4ms, as indicated in Fig. 9.

For signal ITD 0.4 ms, peak prominences are smaller than for the unphysiological $S_\pi N_0$ condition leading to higher detection thresholds and thus smaller BMLDs. However, consistent with the literature [4] the threshold for $S_0N_{0.4}$ is higher than that for $S_{0.4}N_0$.

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 delay model 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 500 Hz, stimulus separation, as measured by Pearson correlation model between output and signal source, was next tested for mixtures of 2 and 3 human speakers without low-pass filtering (Fig. 10). Correlation coefficients between individual speech sounds and model outputs in 20 ms windows revealed high reconstruction acuity

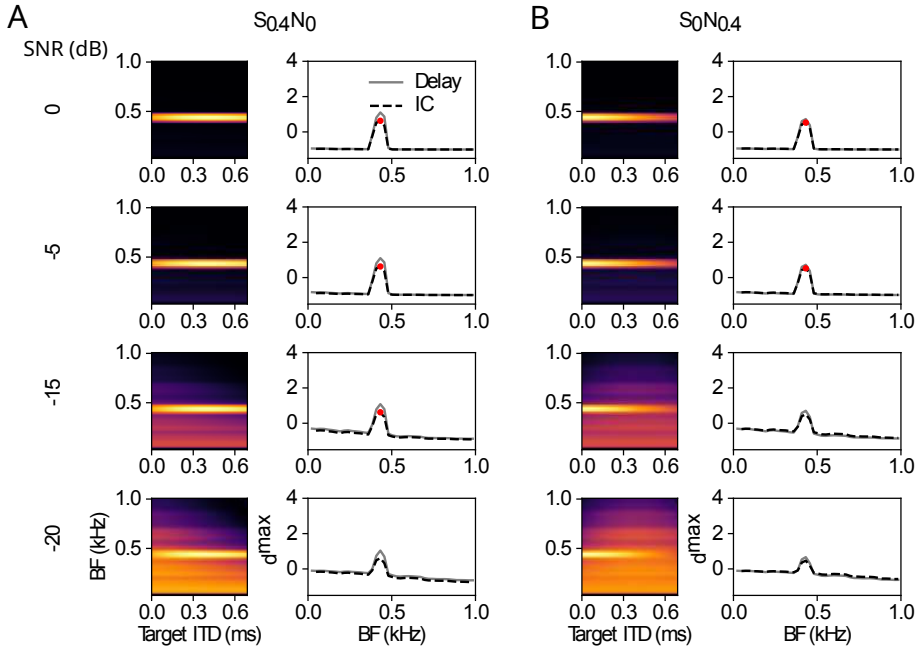

Fig 9. Binaural unmasking with physiological ITDs. Same as in Fig. 7. Note the SNR scale is reduced as compared to Figs. 7 and 8.

in temporally alternating windows relating to one of the speakers (Fig. 10A). Both the IC model and the delay line model exhibited similar pattern, 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. The firing rate pattern across the population of neurons with varying target ITDs reflects the spatial locations in the contralateral hemisphere (Fig. 10B). Sounds with ITDs corresponding to the ipsilateral hemisphere evoke largest rates at target ITD 0. For acoustic verification, audio samples of the reconstructed wave forms, the mixtures, as well as the original sounds can be found as Supporting Information S1.

In a final set of simulations, the model was extended by an adaptive mechanism (see Methods), assuming that the SOC inputs show firing rate adaptation as reported from measurements [66, 67]. Simulation outcomes are shown in Fig. 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 (Fig. 11A), which generates a trend towards balanced LSO and MSO activity (45° angle in the MSO-LSO plane). In sounds with strong amplitude modulations such as speech (Fig. 11C) 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 [68]. Such adaptation-induced fluctuations in neural activation recruits more neurons (i.e., neurons within a broader range of target 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

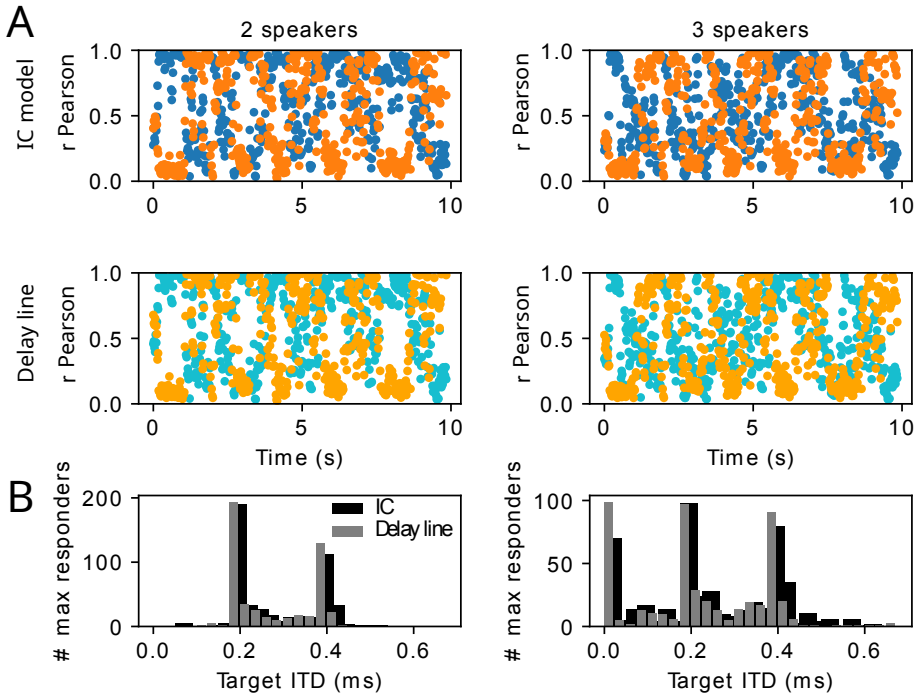

Fig 10. Deciphering a speech scene. Sound scenes were constructed from two (left column: female ITD 0.2ms, male ITD 0.4ms) and three (right column: same as left plus one additional male at ITD -0.2ms) speech signals at physiological ITDs. (A) Pearson's r values were computed between the first two isolated speech sounds (female: blue, male: orange) and the broad band reconstructed IC model output [top row according to Fourier back transform of eq. (11)] and the delay line model [bottom row according to eq. (5)] in snippets of 20ms. Dots represent the mean Pearson's r averaged over all target ITDs (0 to 0.7ms). (B) Histograms of how often a neuron with a given target ITD had the maximal response among its peers (with other target ITD) in a 20ms snippet (IC model: black; Delay line model: grey). Speech signals were taken from a free audiobook.

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 theoretically optimal performance of the 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 [43].

The mammalian ascending auditory pathway leading up to the ITD sensitive IC neurons involves additional connections and nuclei beyond the basic circuit that has inspired our model (Fig. 1). Furthermore, there is species variability [69]. Most importantly, our model omits the dorsal nucleus of the lateral lemniscus (DNLL) [70], which also holds ITD sensitive neurons [71] and projects mostly GABA-ergic inhibition to the IC as well as to its contralateral counterpart. The DNLL has been implicated in echo suppression [72], but direct influence on ITD processing in IC cannot be excluded. We nevertheless refrained from adding these additional pathways to our model to first understand the implications of the most prominent projections in ITD sensitive

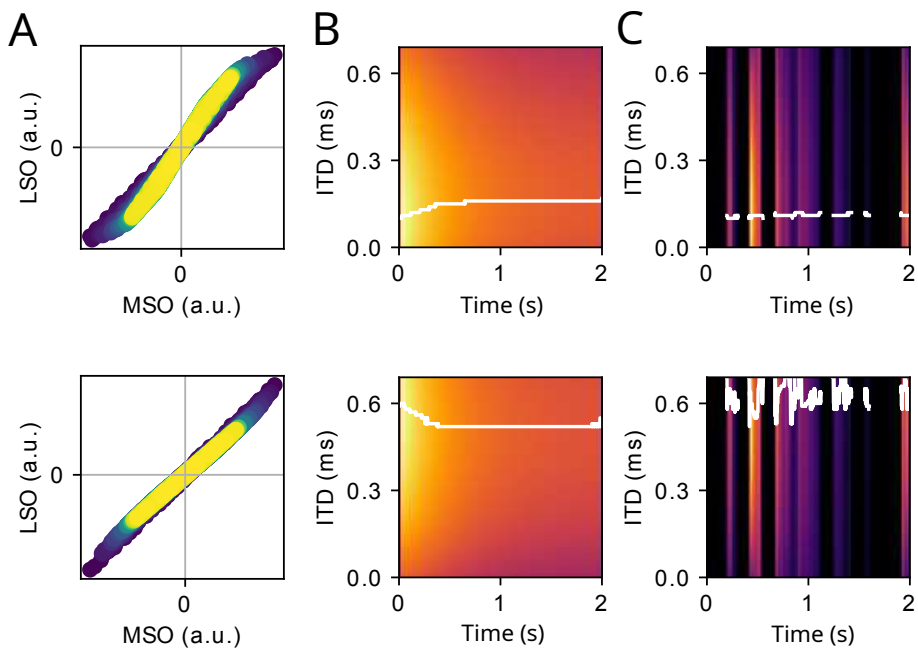

Fig 11. (A) MSO-LSO input plane (see Fig. 4) upon stimulation with a 2 second 400 Hz pure tone with ITD=0.1 ms (top) and ITD=0.6 ms (bottom). The color code reflects the stimulus time (blue early, yellow late). (B) IC population activity (target ITD of IC neuron on y-axis) stimulus ITD=0.1 ms (top) and 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.

mammals, and leave it for future work to understand potential functional consequences of the slower GABAergic DNLL inputs.

Current neurobiological circuit models of binaural hearing have a strong focus on explaining sound localization, including its mechanisms [57, 73–78], encoding [10, 25, 29, 30, 73, 79] and perception [26, 54]. Sound localization and binaural unmasking, however, involve the 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. Our modeling approach also 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 [80] and the brainstem [81] 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 [51]. 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 inhibition) with the original sound wave in the time domain and thus also a direct comparison with the delay line model, which also is formulated in the time domain.

Our model proposes two fitting parameters, a and ψ_c, to be optimized to maximize the firing rate for a single sound stimulus at a given target ITD (under additional constraints that a is not too large, not always zero and predominantly positive). Optimizing a to maximize firing rate in an inhibitorily connected network is conceivable via Hebbian mechanisms in the feed-forward synaptic pathways implementing a winner-take-all type of receptive fields as suggested for visual maps [82]. Optimization of ψ_c does not need to occur on a single cell basis (because the optimal value 0.3 cyc is universal) and therefore could arise from general activity dependent myelin plasticity [83, 84]. The general circuit design of the mammalian auditory pathway, however, would also suggest that part of this optimization could have been achieved through evolution [69].

The present model furthermore resolves a shortcoming of two/four-channel models [22–24, 26, 28, 54], which so far do not provide a conceivable explanation for attentional effects. Mammals are able to focus their attention on a specific sound location of behavioral importance [85–87], however, neural mechanisms of attention typically suggest a bias towards neurons with receptive fields in the attended area [88]. Such a bias can be easily mechanistically implemented by facilitating the response or transmission of neurons with "attended" receptive fields [89], 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. Our model assumptions thus argue 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 [90, 91]. 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 [32, 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 [92]. Also some contemporary bat species have too small heads to use ITDs, but nevertheless possess a monaural MSO [93]. Therefore the general assumption that the evolution 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.

Supporting information

S1 file. Sound files. Zip files contains three original sound snippets `female1.wav`, `male2.wav`, `male3.wav` mixtures of two and three sources `mixture.wav`, `mixture3.wav` and reconstructions according to the delay line model (Jeffress) and the IC model. The original sounds were taken from a free audiobook (https://librivox.org/the-adventures-of-huckleberry-finn-version-5-dramatic-reading-by-

mark-twain/)

Acknowledgments

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.

Code Availability

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

Author Contributions

CL: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Software, Supervision, Visualization, Writing – Original Draft Preparation; SG: Validation, Investigation, Methodology, Writing – Review & Editing

Financial Disclosure

This work was supported by the German Research Association (DFG) CRC 870, TPB01 (to CL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Blauert J, service) SO. The Technology of Binaural Understanding. 1st ed. Modern Acoustics and Signal Processing,. Cham :: Springer International Publishing :; 2020. Available from: `https://doi.org/10.1007/978-3-030-00386-9`.

2. Nachmani E, Adi Y, Wolf L. Voice Separation with an Unknown Number of Multiple Speakers. arXiv. 2020;2003.01531.

3. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. Journal of the acoustical society of America. 1953;25:975–979.

4. van der Heijden M, Trahiotis C. Masking with interaurally delayed stimuli: The use of "internal" delays in binaural detection. The Journal of the Acoustical Society of America. 1999;105(1):388–399. doi:10.1121/1.424628.

5. Bernstein LR, Trahiotis C. Behavioral manifestations of audiometrically-defined "slight" or "hidden" hearing loss revealed by measures of binaural detection. J Acoust Soc Am. 2016;140(5):3540–3548.

6. Grothe B, Pecka M, McAlpine D. Mechanisms of sound localization in mammals. Physiol Rev. 2010;90(3):983–1012.

7. Grothe B, Leibold C, Pecka M. The Medial Superior Olivary Nucleus: Meeting the Need for Speed. In: The Oxford Handbook of the Auditory Brainstem. Oxford University Press; 2019.Available from: `https://doi.org/10.1093/oxfordhb/9780190849061.013.9`.

8. Yin TCT, Smith PH, Joris PX. Neural Mechanisms of Binaural Processing in the Auditory Brainstem. Compr Physiol. 2019;9(4):1503–1575.

9. Wenstrup JJ, Fuzessery ZM, Pollak GD. Binaural neurons in the mustache bat's inferior colliculus. II. Determinants of spatial responses among 60-kHz EI units. J Neurophysiol. 1988;60(4):1384–1404.

10. McAlpine D, Jiang D, Palmer AR. A neural code for low-frequency sound localization in mammals. Nat Neurosci. 2001;4(4):396–401.

11. Joris PX, Van de Sande B, Louage DH, van der Heijden M. Binaural and cochlear disparities. Proc Natl Acad Sci U S A. 2006;103(34):12917–12922.

12. Agapiou JP, McAlpine D. Low-frequency envelope sensitivity produces asymmetric binaural tuning curves. J Neurophysiol. 2008;100(4):2381–2396.

13. McAlpine D, Jiang D, Shackleton TM, Palmer AR. Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus. J Neurosci. 1998;18(15):6026–6039.

14. Fitzpatrick DC, Kuwada S, Batra R. Transformations in processing interaural time differences between the superior olivary complex and inferior colliculus: beyond the Jeffress model. Hear Res. 2002;168(1-2):79–89.

15. Wang L, Devore S, Delgutte B, Colburn HS. Dual sensitivity of inferior colliculus neurons to ITD in the envelopes of high-frequency sounds: experimental and modeling study. J Neurophysiol. 2014;111(1):164–181.

16. Klug A, Bauer EE, Hanson JT, Hurley L, Meitzen J, Pollak GD. Response selectivity for species-specific calls in the inferior colliculus of Mexican free-tailed bats is generated by inhibition. J Neurophysiol. 2002;88(4):1941–1954.

17. Grantham DW, Wightman FL. Detectability of varying interaural temporal differences. J Acoust Soc Am. 1978;63(2):511–523.

18. Yassin L, Pecka M, Kajopoulos J, Gleiss H, Li L, Leibold C, et al. Differences in synaptic and intrinsic properties result in topographic heterogeneity of temporal processing of neurons within the inferior colliculus. Hear Res. 2016;341:79–90.

19. Joris PX, van de Sande B, Recio-Spinoso A, van der Heijden M. Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation. J Neurosci. 2006;26(1):279–289.

20. Siveke I, Ewert SD, Grothe B, Wiegrebe L. Psychophysical and physiological evidence for fast binaural processing. J Neurosci. 2008;28(9):2043–2052.

21. Shackleton TM, Palmer AR. The time course of binaural masking in the inferior colliculus of guinea pig does not account for binaural sluggishness. J Neurophysiol. 2010;104(1):189–199.

22. Hancock KE, Delgutte B. A physiologically based model of interaural time difference discrimination. J Neurosci. 2004;24(32):7110–7117.

23. MARQUARDT T, MCALPINE D. 44 A p-Limit for Coding ITDs: Implications for Binaural Models. Hearing-From Sensory Processing to Perception. 2007; p. 407.

24. Lesica NA, Lingner A, Grothe B. Population coding of interaural time differences in gerbils and barn owls. J Neurosci. 2010;30(35):11696–11702.

25. Lüling H, Siveke I, Grothe B, Leibold C. Frequency-invariant representation of interaural time differences in mammals. PLoS Comput Biol. 2011;7(3):e1002013.

26. Lingner A, Pecka M, Leibold C, Grothe B. A novel concept for dynamic adjustment of auditory space. Sci Rep. 2018;8(1):8335.

27. Groß S. Sound source coding in the azimuthal plane: separating sounds via short-term interaural time difference estimations. Ludwig-Maximilians-Universität München; 2021. Available from: `http://nbn-resolving.de/urn:nbn:de:bvb:19-301024`.

28. Encke J, Dietz M. A hemispheric two-channel code accounts for binaural unmasking in humans. Commun Biol. 2022;5(1):1122.

29. Harper NS, McAlpine D. Optimal neural population coding of an auditory spatial cue. Nature. 2004;430(7000):682–686.

30. Goodman DF, Benichoux V, Brette R. Decoding neural responses to temporal cues for sound localization. Elife. 2013;2:e01312.

31. Hirsh IJ. The influence of interaural phase on interaural summation and inhibition. The Journal of the Acoustical Society of America. 1948;20(4):536–544.

32. Culling JF, Lavandier M. Binaural Unmasking and Spatial Release from Masking. In: Litovsky RY, Goupell MJ, Fay RR, Popper AN, editors. Binaural Hearing: With 93 Illustrations. Cham: Springer International Publishing; 2021. p. 209–241. Available from: `https://doi.org/10.1007/978-3-030-57100-9_8`.

33. van der Heijden M, Joris PX. Interaural correlation fails to account for detection in a classic binaural task: dynamic ITDs dominate N0Spi detection. J Assoc Res Otolaryngol. 2010;11(1):113–131.

34. Dietz M, Ewert SD, Hohmann V, Kollmeier B. Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences. Brain Research. 2008;1220:234–245. doi:https://doi.org/10.1016/j.brainres.2007.09.026.

35. Dietz M, Ashida G. Computational Models of Binaural Processing. In: Litovsky RY, Goupell MJ, Fay RR, Popper AN, editors. Binaural Hearing: With 93 Illustrations. Cham: Springer International Publishing; 2021. p. 281–315. Available from: `https://doi.org/10.1007/978-3-030-57100-9_10`.

36. Bernstein LR, Trahiotis C. The normalized correlation: Accounting for binaural detection across center frequency. The Journal of the Acoustical Society of America. 1996;100(6):3774–3784. doi:10.1121/1.417237.

37. van de Par S, Kohlrausch A. A new approach to comparing binaural masking level differences at low and high frequencies. J Acoust Soc Am. 1997;101(3):1671–1680.

38. Beutelmann R, Brand T. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2006;120(1):331–342.

39. Wan R, Durlach NI, Colburn HS. Application of a short-time version of the Equalization-Cancellation model to speech intelligibility experiments with speech maskers. J Acoust Soc Am. 2014;136(2):768–776. doi:10.1121/1.4884767.

40. Biberger T, Ewert SD. Towards a simplified and generalized monaural and binaural auditory model for psychoacoustics and speech intelligibility. Acta Acust. 2022;6:23. doi:10.1051/aacus/2022018.

41. Stern RM, Trahiotis C. Models of binaural interaction. Handbook of perception and cognition. 1995;6:347–386.

42. Penrose R. A generalized inverse for matrices. Mathematical Proceedings of the Cambridge Philosophical Society. 1955;51(3):406–413. doi:10.1017/S0305004100030401.

43. Karino S, Smith PH, Yin TC, Joris PX. Axonal branching patterns as sources of delay in the mammalian auditory brainstem: a re-examination. J Neurosci. 2011;31(8):3016–3031.

44. Kapfer C, Seidl AH, Schweizer H, Grothe B. Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. Nat Neurosci. 2002;5(3):247–253.

45. Couchman K, Grothe B, Felmy F. Medial superior olivary neurons receive surprisingly few excitatory and inhibitory inputs with balanced strength and short-term dynamics. J Neurosci. 2010;30(50):17111–17121.

46. Roberts MT, Seeman SC, Golding NL. A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry. Neuron. 2013;78(5):923–935.

47. Park TJ, Grothe B, Pollak GD, Schuller G, Koch U. Neural delays shape selectivity to interaural intensity differences in the lateral superior olive. J Neurosci. 1996;16(20):6554–6566.

48. Tollin DJ, Yin TC. Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive. J Neurosci. 2005;25(46):10648–10657.

49. Beiderbeck B, Myoga MH, Müller NIC, Callan AR, Friauf E, Grothe B, et al. Precisely timed inhibition facilitates action potential firing for spatial coding in the auditory brainstem. Nat Commun. 2018;9(1):1771.

50. Franken TP, Bondy BJ, Haimes DB, Goldwyn JH, Golding NL, Smith PH, et al. Glycinergic axonal inhibition subserves acute spatial sensitivity to sudden increases in sound intensity. eLife. 2021;10:e62183. doi:10.7554/eLife.62183.

51. Callan AR, Heß M, Felmy F, Leibold C. Arrangement of Excitatory Synaptic Inputs on Dendrites of the Medial Superior Olive. J Neurosci. 2021;41(2):269–283.

52. Glendenning KK, Baker BN, Hutson KA, Masterton RB. Acoustic chiasm V: inhibition and excitation in the ipsilateral and contralateral projections of LSO. J Comp Neurol. 1992;319(1):100–122.

53. Brückner S, Rübsamen R. Binaural response characteristics in isofrequency sheets of the gerbil inferior colliculus. Hear Res. 1995;86(1-2):1–14.

54. Shackleton TM, Skottun BC, Arnott RH, Palmer AR. Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of Guinea pigs. J Neurosci. 2003;23(2):716–724.

55. Furst M, Aharonson V, Levine RA, Fullerton BC, Tadmor R, Pratt H, et al. Sound lateralization and interaural discrimination. Hear Res. 2000;143(1-2):29–42.

56. Litovsky RY, Fligor BJ, Tramo MJ. Functional role of the human inferior colliculus in binaural hearing. Hear Res. 2002;165(1-2):177–188.

57. Brand A, Behrend O, Marquardt T, McAlpine D, Grothe B. Precise inhibition is essential for microsecond interaural time difference coding. Nature. 2002;417(6888):543–547.

58. Grothe B, Park TJ. Time can be traded for intensity in the lower auditory system. Naturwissenschaften. 1995;82(11):521–523.

59. McAlpine D, Jiang D, Palmer AR. Interaural delay sensitivity and the classification of low best-frequency binaural responses in the inferior colliculus of the guinea pig. Hear Res. 1996;97(1-2):136–152.

60. van de Par S, Kohlrausch A. Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. J Acoust Soc Am. 1999;106(4):1940–1947.

61. Jiang D, McAlpine D, Palmer AR. Responses of neurons in the inferior colliculus to binaural masking level difference stimuli measured by rate-versus-level functions. J Neurophysiol. 1997;77(6):3085–3106.

62. Jiang D, McAlpine D, Palmer AR. Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. J Neurosci. 1997;17(23):9331–9339.

63. Langford TL, Jeffress LA. Effect of Noise Crosscorrelation on Binaural Signal Detection. The Journal of the Acoustical Society of America. 1964;36(8):1455–1458. doi:10.1121/1.1919224.

64. See Supplemental Material at [URL will be inserted by publisher] for reconstructed sounds;.

65. See Supplemental Material at [URL will be inserted by publisher] for sound mixtures;.

66. Magnusson AK, Park TJ, Pecka M, Grothe B, Koch U. Retrograde GABA signaling adjusts sound localization by balancing excitation and inhibition in the brainstem. Neuron. 2008;59(1):125–137.

67. Stange A, Myoga MH, Lingner A, Ford MC, Alexandrova O, Felmy F, et al. Adaptation in sound localization: from GABA(B) receptor-mediated synaptic modulation to perception. Nat Neurosci. 2013;16(12):1840–1847.

68. Dietz M, Marquardt T, Salminen NH, McAlpine D. Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. Proc Natl Acad Sci U S A. 2013;110(37):15151–15156.

69. Williams IR, Ryugo DK. Bilateral and symmetric glycinergic and glutamatergic projections from the LSO to the IC in the CBA/CaH mouse. Front Neural Circuits. 2024;18:1430598.

70. Felmy F. The nuclei of the lateral lemniscus. The Oxford handbook of the auditory brainstem. 2019; p. 445–472. doi:10.1093/oxfordhb/9780190849061.001.0001.

71. Siveke I, Pecka M, Seidl AH, Baudoux S, Grothe B. Binaural response properties of low-frequency neurons in the gerbil dorsal nucleus of the lateral lemniscus. J Neurophysiol. 2006;96(3):1425–1440.

72. Pecka M, Zahn TP, Saunier-Rebori B, Siveke I, Felmy F, Wiegrebe L, et al. Inhibiting the inhibition: a neuronal network for sound localization in reverberant environments. J Neurosci. 2007;27(7):1782–1790.

73. Jeffress LA. A place theory of sound localization. J Comp Physiol Psychol. 1948;41(1):35–39.

74. Leibold C, van Hemmen JL. Spiking neurons learning phase delays: how mammals may develop auditory time-difference sensitivity. Phys Rev Lett. 2005;94(16):168102.

75. Zhou Y, Carney LH, Colburn HS. 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. 2005;25(12):3046–3058.

76. Jercog PE, Svirskis G, Kotak VC, Sanes DH, Rinzel J. Asymmetric excitatory synaptic dynamics underlie interaural time difference processing in the auditory system. PLoS Biol. 2010;8(6):e1000406.

77. Leibold C. Influence of inhibitory synaptic kinetics on the interaural time difference sensitivity in a linear model of binaural coincidence detection. J Acoust Soc Am. 2010;127(2):931–942.

78. Lehnert S, Ford MC, Alexandrova O, Hellmundt F, Felmy F, Grothe B, et al. Action potential generation in an anatomically constrained model of medial superior olive axons. J Neurosci. 2014;34(15):5370–5384.

79. von Bekesy G. Zur Theorie des Hörens; über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schallwirkungen. Physik Zeitschr. 1930;31:824–835, 857–868.

80. Wen B, Wang GI, Dean I, Delgutte B. Dynamic range adaptation to sound level statistics in the auditory nerve. J Neurosci. 2009;29(44):13797–13808.

81. Willmore BDB, King AJ. Adaptation in auditory processing. Physiol Rev. 2023;103(2):1025–1058.

82. von der Malsburg C. Self-organization of orientation sensitive cells in the striate cortex. Kybernetik. 1973;14(2):85–100.

83. Sinclair JL, Fischl MJ, Alexandrova O, Heβ M, Grothe B, Leibold C, et al. Sound-evoked activity influences myelination of brainstem axons in the trapezoid body. J Neurosci. 2017;37(34):8239–8255.

84. Stancu M, Wohlfrom H, Heß M, Grothe B, Leibold C, Kopp-Scheinpflug C. Ambient sound stimulation tunes axonal conduction velocity by regulating radial growth of myelin on an individual, axon-by-axon basis. Proc Natl Acad Sci U S A. 2024;121(11):e2316439121.

85. Rhodes G. Auditory attention and the representation of spatial information. Percept Psychophys. 1987;42(1):1–14.

86. Lee CC, Middlebrooks JC. Auditory cortex spatial sensitivity sharpens during task performance. Nat Neurosci. 2011;14(1):108–114.

87. Myoga MH, Amaro D, Kello V, Gumbert M, Leibold C, Pecka M, et al. The Dynamics of Context-Dependent Space Representations in Auditory Cortex. bioRxiv. 2023;doi:10.1101/2023.11.25.568638.

88. Borji A, Itti L. Optimal attentional modulation of a neural population. Front Comput Neurosci. 2014;8:34.

89. Reynolds JH, Heeger DJ. The normalization model of attention. Neuron. 2009;61(2):168–185.

90. Smith PH, Joris PX, Carney LH, Yin TC. Projections of physiologically characterized globular bushy cell axons from the cochlear nucleus of the cat. J Comp Neurol. 1991;304(3):387–407.

91. Joris PX, Yin TC. Responses to amplitude-modulated tones in the auditory nerve of the cat. J Acoust Soc Am. 1992;91(1):215–232.

92. Grothe B, Pecka M. The natural history of sound localization in mammals – a story of neuronal inhibition. Front Neural Circuits. 2014;8:116.

93. Grothe B, Covey E, Casseday JH. Medial Superior Olive of the Big Brown Bat: Neuronal Responses to Pure Tones, Amplitude Modulations, and Pulse Trains. Journal of Neurophysiology. 2001;86(5):2219–2230. doi:10.1152/jn.2001.86.5.2219.