Executive Summary of the Thesis

Why can you kill mosquitoes in the dark?
Towards solving sound localization with a biologically plausible spiking neural simulation

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Paolo Marzolo

Advisor: Prof. Alberto Antonietti

Co-advisor: Eng. Francesco De Santis

Academic year: 2024-2025

1. Introduction

The task of sound localization consists of identifying the source position of a sound. Differently from sight and somatosensation (touch), stimulus location is not directly linked to a sensory organ region. Hence, sound source position must be *derived* from the original input; this derivation is based on three main features (*auditory cues*): the differences in time and loudness (*level*) between when a sound reaches each ear, namely the interaural time difference (ITD) and interaural level difference (ILD), and the difference in how a sound gets distorted by the obstacles it encounters along its path (the body, the head and the outer ear), depending on its source location. This work focuses on horizontal (azimuthal) localization, for which only the first two cues are relevant. In the mammalian brain, these cues are generally associated with distinct nuclei in the superior olivary complex (SOC), which respectively show increased sensibility to a single cue: the medial superior olive (MSO) for the ITD, and the lateral superior olive (LSO) for the ILD. Unlike the processing of ILD, the processing of ITDs is less understood. Multi-

ple strategies have been proposed; the most impactful is a model presented by Jeffress in 1948. This model assumes that ITDs are processed by an array of simple coincidence detectors (neurons that respond maximally when the two inputs arrive simultaneously), wired with progressively longer delay lines (based on axonal length) that generate internal delays. Then, the coincidence detector that fires maximally will represent the current ITD, generating an internal map of ITDs. The two structures that power the Jeffress scheme (delay lines and ITD map) were confirmed to exist in the avian brain [3]. Due to parallel evolution, it is possible that findings obtained in different lineages (birds) may not exist in the mammalian brainstem, and indeed, there is very weak evidence of these structures in mammals. This has pushed research into looking for the alternative mechanism in use in mammals. One such proposal is Myoga's [4], who focused on the timing of excitatory and inhibitory inputs. The MSO receives four inputs from lower-level nuclei: a pair of inhibition and excitation from each side. The authors found that both sources of inhibition to the MSO can be used to tune its ability to perform coinci-

dence detection, hence changing the ITD value for which the MSO cells will have the maximum response.

The first neural step of the auditory pathway are the auditory nerve fibers (ANFs), which receive the signal after it has been elaborated and transduced into a graded potential by the peripheral processing sections. The source transducers (which are the inner hair cells inside the cochlea) only respond to a specific frequency, called the characteristic frequency (CF): this causes the entire neural population to be organized by frequency. The ANFs then converge to two main populations, the spherical bushy cells (SBCs) and globular bushy cells (GBCs). The SBCs are the main source of excitation to the SOC nuclei, while the GBCs, after an additional step to transform the signal from excitatory to inhibitory, provide the main source of inhibition. The LSO receives excitation from the SBCs and inhibition from the contralateral GBCs (mediated by cells of the medial nucleus of the trapezoid body, MNTB); the MSO receives excitation from both sides of SBCs, and inhibition from ipsilateral GBCs (mediated by cells of the lateral nucleus of the trapezoid body, LNTB) and from contralateral GBCs (once again mediated by cells in the MNTB). We show the complete network, together with the connectivity patterns, the delays and synaptic weights we used, in Figure 1.

The LSO and MSO then converge to the inferior colliculus (IC). Although the IC comprises three regions - central nucleus, dorsal cortex, and external cortex - the central nucleus (ICC) is the primary region of interest for auditory processing, together with the brachium of the IC. The ICC receives input from both of the SOC nuclei discussed so far, LSO and MSO, in addition to other sources; specifically, each IC receives input from ipsilateral MSO and contralateral LSO. These synapses are mostly excitatory, forming an EE scheme (excitatory-excitatory) [3]. Cells in the ICC have shown varying properties and classified accordingly: type V units, sensitive to ITDs, type I units, sensitive to ILDs, and type IV units, resembling neurons in the dorsal cochlea nucleus, which processes spectral cues. Evidence for a spatial map has not been found, and some results [6] suggest that auditory cues

may remain more segregated than previously thought. Overall, the literature does not definitively answer whether and how auditory cues are integrated above the SOC.

In this work, we create a bioplausible network for auditory localization, with three main steps. First, we mimic biological auditory processing structures to create realistic inputs for a neural simulation, containing both intermediate populations and the two nuclei of interest. Second, we use Myoga's principles to create a plausible MSO that shows increased activity for contralateral sounds, as seen in literature. Finally, we investigate whether this MSO adds information to the LSO for the localization of tones at different frequencies, by creating a synthetic inferior colliculus (the anatomical site of convergence of LSO and MSO) with a simple EE scheme and testing its performance.

2. Objectives

Given the state of research around sound localization, this work improves an existing, simplified, neural-only computational model [5], with three main objectives:

1. Improve the overall bioplausibility of the existing model by:
   (a) Implementing peripheral sections at three levels of bioplausibility: (1) *Pulse packet* (2) *Gammatone-based* (3) *Tan-Carney realistic mammalian cochlea*
   (b) Evaluating the effect of these peripheral sections on SOC nuclei (LSO and MSO)
2. Use Myoga's principles to produce a contralateral-favoring MSO without the use of dedicated delay lines
3. Consider an EE (contralateral LSO and ipsilateral MSO) scheme for auditory cue integration above the SOC with a synthesized IC model and quantitatively evaluate it on three metrics:
   (a) *Zero-degree accuracy*: Since behavioral results show no difference between right and left azimuthal localization accuracy, the center point (zero degrees) should show equal activation in both populations;
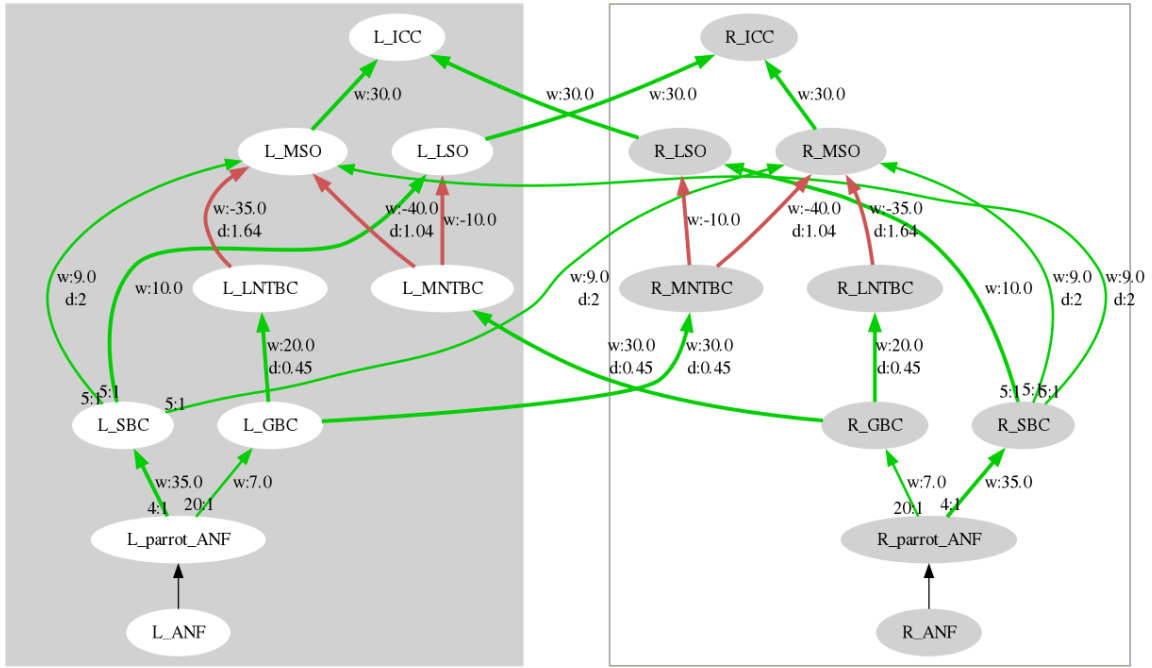   (b) *Increased sensitivity around zero*: Just

Figure 1: The complete network, from ANFs to IC. Green connections are excitatory, red are inhibitory. Next to each connection, its synaptic weight and delay. From ANFs and bushy cells, the notation X:Y shows convergence information.

noticeable difference, a metric that shows the lowest detectable degree difference between sounds, is lowest in humans around zero azimuth. Hence, a steeper curve is an improvement;

(c) *Range*: Difference between maximum and minimum total population spike index to maximize the difference in lateralization.

3. Methods

This project was split in two sections: the first focused on increasing the bioplausibility of the peripheral system, and the second one implemented the neural processing. The modeling of the peripheral section, from the sound to the spiking output of ANFs, was implemented using Brian2Hears [1], an extension to Brian 2 [8], a widely used open-source spiking neural network simulator. The modeling of the central nervous system was developed using NEST [2], another well known simulator.

3.1. Peripheral section

Using Brian 2 Hears, we implemented two peripheral sections at different levels of bioplausibility: a simple gammatone model and the more advanced Tan-Carney model [9]. Figure 2 shows the pathway common to all peripheral models. The first step is modeling the compound effect of ILD, ITD and spectral cues, to differentiate how sound reaches the acoustic canal of each ear. This compound effect for how a listener's environment and body composition affect sound arrival times, level, and spectral content is called the Head Related Transfer Function (HRTF), a digital filter. Once a sound has been filtered through an HRTF, the signal arriving to each ear goes through peripheral processing, which varies between the different models we tested, and causes an ANF spiking pattern. The ANF spike trains are the final result of the peripheral processing pipeline. We also compared these peripheral processing pipelines to a simpler alternative solution, based on pulse packet generators, which produce a spike train containing Gaussian-distributed spikes centered about given times. A Gaussian pulse packet is a given number of spikes with normal distributed random displacements from the center time of the pulse.
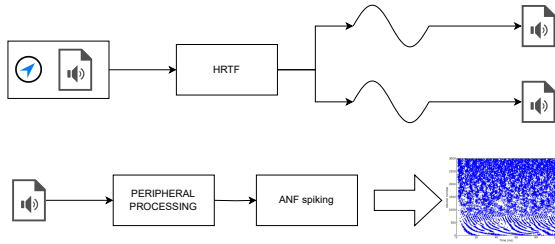
Figure 2: Block scheme for auditory peripheral processing

3.2. Neural processing

We recorded the same metrics for all neural layers. As we need to compare results among different trials, to verify whether the network is able to distinguish between sounds coming from different angles, we used a metric that could express the response of the entire population. We use the total population spike rate: all spikes from each population are tallied and scaled to the duration of the stimulus. Comparisons between different populations are not very meaningful, as due to convergence different populations have different neuron counts. In addition, to evaluate how cells at different CFs were contributing to the overall spiking pattern, we included vertical histograms for each angle.

3.3. IC

Our last objective is to evaluate a possible way to combine cue processing by the LSO and the MSO. We draw a simple excitatory connection from both of them (the IC receives inputs from ipsilateral MSO and contralateral LSO), and we evaluate if the result improves compared to the LSO alone. To measure the *improvement*, we computed the difference of the total population spikes between the two ICs for each azimuthal location. In the literature, the majority of IC neurons show ILD sensitivity, which would suggest a crucial role for the LSO in determining IC responses [3], and ILDs are the strongest cue for spatial tuning in the IC [7]. Because of this, we fit the difference in population activation to a LSO-like rate curve, described by a sigmoid with the following equation:

$$\frac{L}{1 + e^{-k(x-x_0)}} + b$$

We then define three metrics which will be the values that minimize non-linear least squares
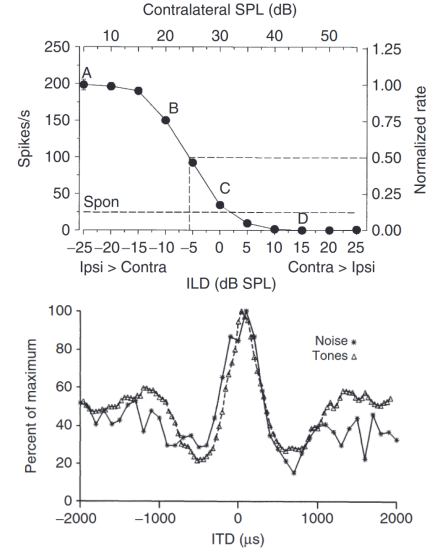

Figure 3: The response of SOC nuclei to the cue they are most sensitive to, adapted from [10]. In the top plot, the response of an LSO cell to increasing contralateral level (increasing ILD). In the bottom plot, the response of an MSO cell to varying ITD. Notice the marked ipsilateral preference for the LSO, and contralateral-favoring response of the MSO.

when applying the sigmoid to our data. The correspondence is as follows:

1. x_0: Closeness of center point to zero (to be minimized);
2. k: Steepness around zero (to be maximized);
3. $L - b$: Range, or the difference between maximum and minimum total population spike index (to be maximized).

In addition, the R^2 value will tell us how well the difference in IC spike counts can be fit to the sigmoid.

4. Results

4.1. Peripheral section

We evaluated peripheral processing pathways based on the ANF response, focusing on vector strength (VS), a measure of phase synchrony, or how well spike timings are synchronized to a specific phase in each period. Some of our results are collected in Figure 4, where we plot the phase responses, used to compute the VS, of a single ANF, one of the ten connected to the IHC. The selected ANF has an illustrative characteristic frequency (i.e., 500 Hz), receiving three-second
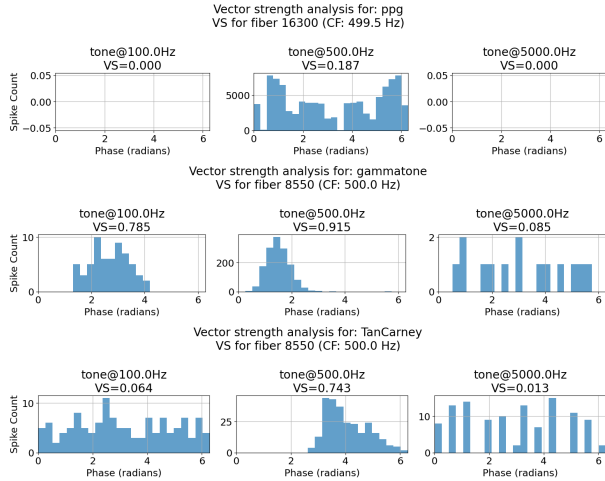
Figure 4: Each plot shows the vector strength and phase response of the same ANF in different peripheral processing pathways (in rows) and to different inputs (in columns, 100 Hz, 500 Hz, and 5 kHz).

tones with frequency 100 Hz, 500 Hz, and 5 kHz. The results show that using more biologically plausible peripheral processing results in higher-quality ANF spike trains: (1) pulse packet generators were extremely faithful in tonotopic activation, as the only spiking ANFs are those in the immediate surrounding (a range of 2100 ANFs of the total 3500), which is in disagreement with both the gammatone-based and the Tan-Carney processing pathways (2) the phase locking exhibited by the synthetic processing was very limited (VS = 0.187) (3) the gammatone-based processing shows very accurate phase locking (VS = 0.915) for ANFs whose CFs were close to the input frequency, with a slow falloff for input frequencies further away from the CF (e.g., VS = 0.785 for a 100 Hz tone) (4) the total number of spikes with gammatone-based processing is much higher than the most bioplausible version at CF, but falls off very quickly, possibly masking a lower phase locking than the one shown here (5) the Tan-Carney-based processing shows a moderate VS for 500 Hz input (VS = 0.743) and lower spiking rates at CF, with respect to the gammatone (6) in the Tan–Carney model, the 100 Hz input activates quite strongly, even if the CF is further away. These findings show that both the gammatone-based and the Tan-Carney models present a sufficient level of bioplausibility for further processing. In addition, the Tan-Carney model shows higher

vector strength (hence phase locking), more realistic spike counts with a less drastic falloff and some spontaneous rates, and more realistic falloff of ANF activation as the input tone frequency changes.

4.2. Higher centers

LSO   In our testing, the behavior of the LSO was remarkably stable. Even when receiving input tones at low frequencies, for which the ILD is small as the attenuation given by the head is negligible, the sigmoidal shapes of left and right LSOs were unmistakable and very resistant to perturbations. The LSO remained consistent when we tried to change membrane time constants and connection weights. The typical dual sigmoidal response (see Fig3) was also generated equally well by both realistic cochleas, as shown in 5. The difference between spike numbers of the two sides increased as frequency increased, as is to be expected with the rising ILD. When using the gammatone-based peripheral section, the LSO failed above 10 kHz. The vertical histograms show an additional difference between results: when using the gammatone-based peripheral section, the portion of population spiking (and contributing to the result) is very focused around the fibers with CF close to the input; instead, the Tan-Carney based results show more uniform activation of the entire population, with an increased spike rate for lower-frequency neurons.

MSO   As with the LSO, the largest difference between the two bioplausible peripheral processing models is the extent of involved frequency ranges: while the gammatone-based processing only activates a narrow portion of ANFs (and hence the rest of the network), the Tan-Carney processing generates a broader activation. At the same time, MSO cells with high CF show a limited activation, possibly due to the slower decay time of inhibition, which, at high frequencies, stops cells almost completely. Decay with rising frequency is not the same between the two input sources: while the gammatone-based falls off quickly, losing differentiating abilities at 800 Hz, the Tan-Carney-based shows higher spiking rates at contralateral angles also at 10 kHz.

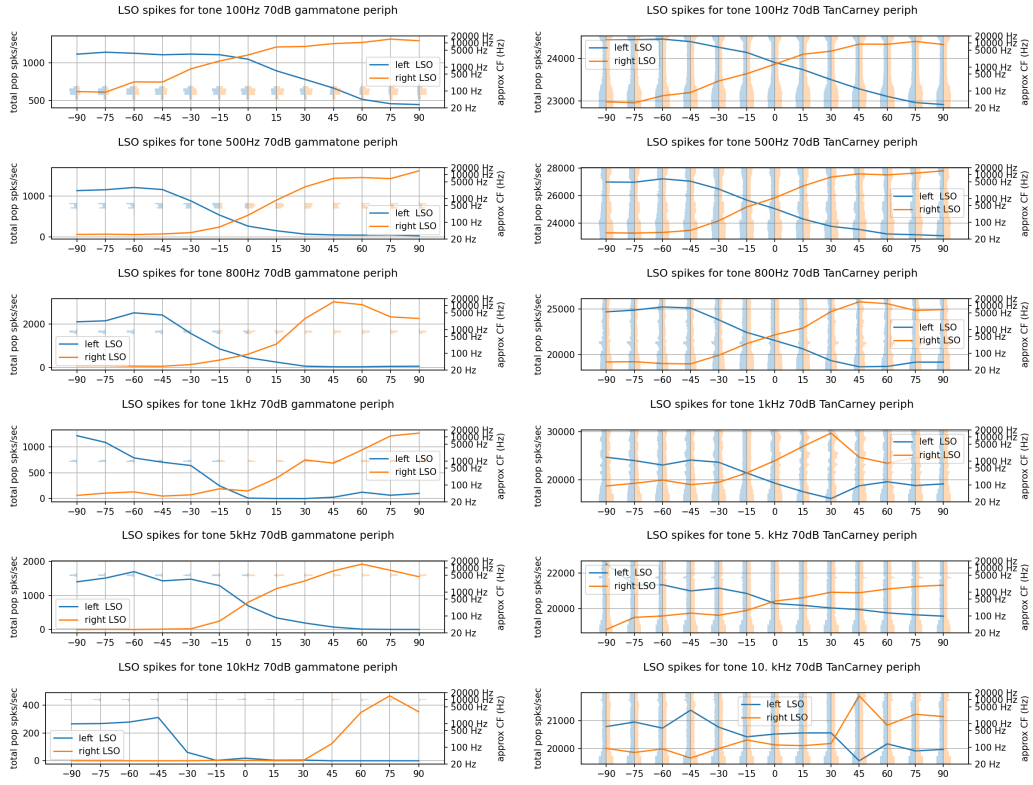Although the impact of the contralateral pref-

Figure 5: Population spike rates for the LSO with growing frequencies, for gammatone-based (left) and Tan-Carney (right) processing.
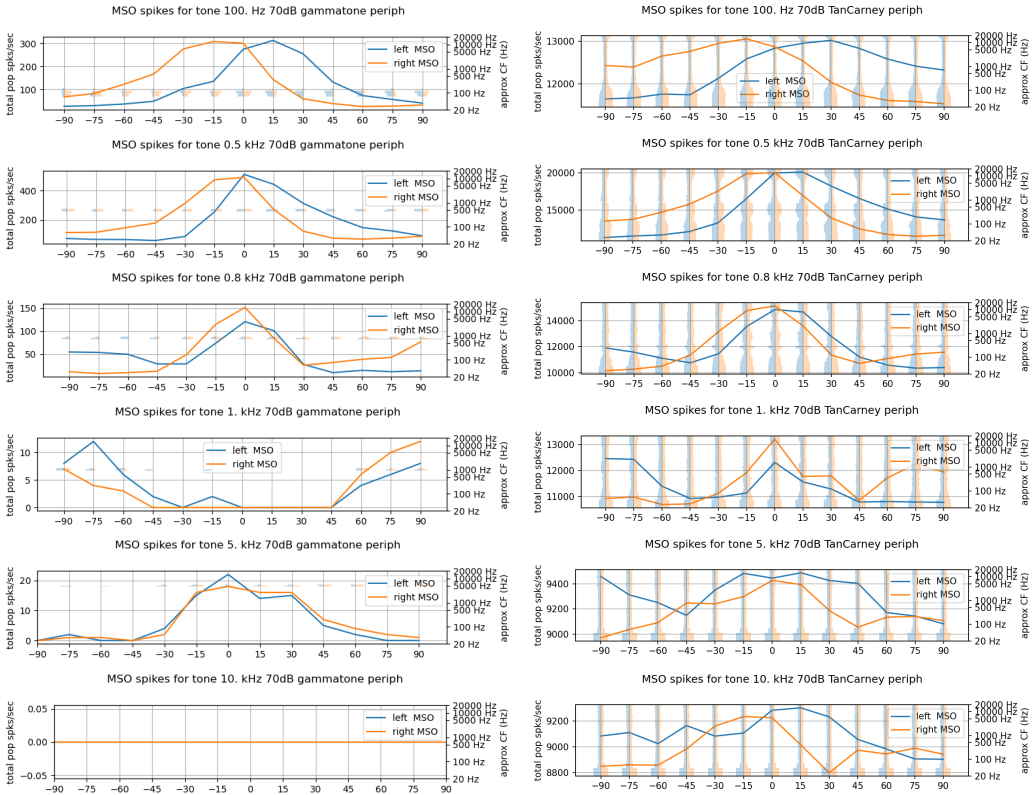

Figure 6: Population spike rates for the MSO with growing frequencies, for gammatone-based (left) and Tan-Carney (right) processing.

erence is limited, MSO behavior shows that Myoga's principles to influence coincidence detection are applicable. This warrants additional research, focused on verifying whether it is possible to fully replicate MSO recordings in larger animals using differences in arrival times of inhibition and excitation, without dedicated delay lines or other, less bioplausible interactions.

4.3. Inferior Colliculus

Finally, we consider whether the integration of cues from MSO and LSO proved beneficial to the IC. To determine the impact of the MSO on the IC curves, we focus on the most bioplausible version of our peripheral processing (i.e., the Tan-Carney model). In Figure 7, we plot the difference of total population spikes between the right and left LSO (top) and IC (bottom). Note that the IC integrates the input of the contralateral LSO and ipsilateral MSO. The results are not particularly striking: while the steepness around the center point improved, this came at a cost for range, which decreased in the IC compared to the LSO. Overall, the zero crossing accuracy was the most significant improvement, which at low frequencies was improved significantly. At the same time, due to database constraints, our data points are 15 degrees apart, and even the largest improvement in zero crossing was by around 3 degrees.

5. Conclusions

With this work, we have shown evidence for three claims: (1) bioplausible peripheral processing achieves ANF spiking patterns that contain fundamental information for localization (as seen with the phase response); (2) using Myoga's principles for a bioplausible MSO results in a contralateral-favoring MSO; (3) we were unable to find a significant improvement in the IC response curves when using a simple excitatory-excitatory connection scheme with the SOC nuclei.

References

[1] Bertrand Fontaine, Dan F. M. Goodman, Victor Benichoux, and Romain Brette. Brian hears: Online auditory processing using vectorization over channels. 5. Publisher: Frontiers Library: https://github.com/brian-team/brian2hears/.

[2] Marc-Oliver Gewaltig and Markus Diesmann. NEST (NEural simulation tool). 2(4):1430.

[3] Benedikt Grothe, Michael Pecka, and David McAlpine. Mechanisms of sound localization in mammals. 90(3):983–1012.

[4] Michael H. Myoga, Simon Lehnert, Christian Leibold, Felix Felmy, and Benedikt Grothe. Glycinergic inhibition tunes coincidence detection in the auditory brainstem. 5(1):3790. Publisher: Nature Publishing Group.

[5] Francesco De Santis. A computational model of the mammalian brainstem to solve sound localization.

[6] Sean J. Slee and Eric D. Young. Alignment of sound localization cues in the nucleus of the brachium of the inferior colliculus. 111(12):2624–2633. Publisher: American Physiological Society.

[7] Sean J. Slee and Eric D. Young. Linear processing of interaural level difference underlies spatial tuning in the nucleus of the brachium of the inferior colliculus. 33(9):3891–3904.

[8] Marcel Stimberg, Romain Brette, and Dan FM Goodman. Brian 2, an intuitive and efficient neural simulator. 8:e47314. Publisher: eLife Sciences Publications, Ltd.

[9] Qing Tan and Laurel H. Carney. A phenomenological model for the responses of auditory-nerve fibers. II. nonlinear tuning with a frequency glide. 114(4):2007–2020.

[10] Tom C.T. Yin, Phil H. Smith, and Philip X. Joris. Neural mechanisms of binaural processing in the auditory brainstem. In Ronald Terjung, editor, *Comprehensive Physiology*, pages 1503–1575. Wiley, 1 edition. tex.ids= yinNeuralMechanismsBinaural2019a, yinNeuralMechanismsBinaural2019b.
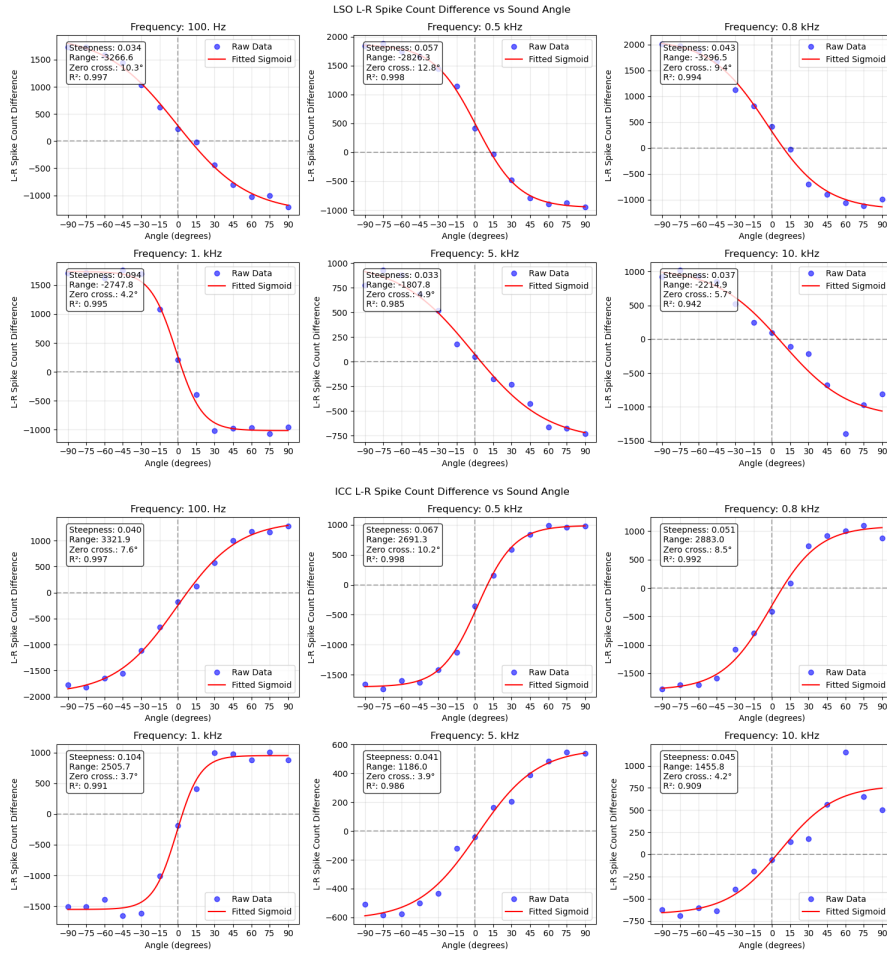
Figure 7: Comparison of sigmoid fit for all frequencies. Top: LSO. Bottom: IC. Because the contralateral LSO connects to IC, its slope is in the opposite direction.