POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Why can you kill mosquitoes in the dark?

Towards solving sound localization with a biologically plausible spiking neural simulation

Tesi di Laurea Magistrale in
Computer Science and Engineering - Ingegneria Informatica

Paolo Marzolo, 10668259

Advisor:
Prof. Alberto Antonietti

Co-advisors:
Eng. Francesco De Santis

Academic year:
2024-2025

Abstract: The ability to localize sounds in space is one of the most studied aspects of hearing. Yet, despite the apparent simplicity, the exact mechanisms behind sound localization in mammals are still uncertain. In this thesis we started from an existing spiking neuronal model mimicking the auditory brainstem circuitry and its tonotopic organization and we increased its bioplausibility. We applied a dual focus on inputs and outputs of the network: using advanced peripheral processing models, we simulated how sound is received and transduced into neural signals by the cochlea; these signals are processed by a neural simulation and integrated by higher nuclei. We then investigated the most recent proposals for this integration and explored a possible way to achieve further processing in the midbrain.

Key-words: sound localization, computational neuroscience, neural simulation

Contents

1. Background

1.1. Introduction

The task of sound localization consists of identifying the source position of a sound. This is possible from hearing alone. Positions are expressed using two angles: the azimuth, the angle on the horizontal plane, and the elevation, the angle on the vertical plane 1. 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 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.

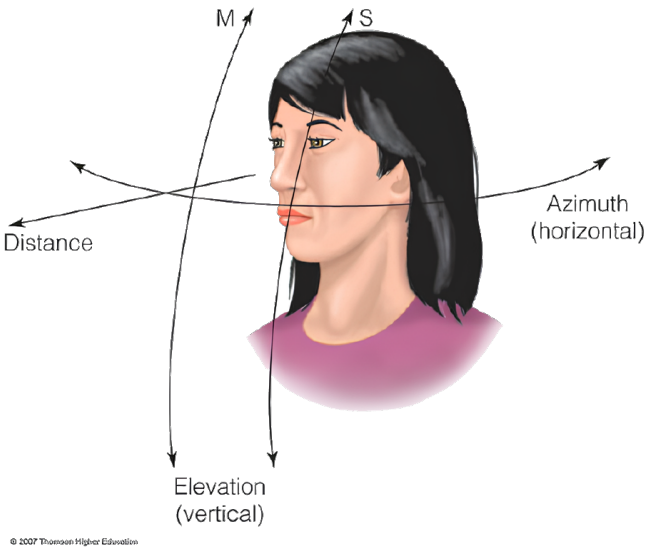


Figure 1: How a sound source location is expressed: azimuthal angle, elevation angle and distance

This task is essential and complex under multiple perspectives, which we will briefly outline here. Sound localization is fundamental for survival, whether predator or prey; it is a foundational part of extracting relevant sounds from a noisy environment; as such, it is common to many different species. And yet, although its usefulness is undeniable and we never question our ability to localize sounds, detecting and localizing air-borne sounds needs specialized physiological structures. When transmitted inside the ear, airwaves become fluid waves, which have a much higher acoustic impedance (i.e., how hard it is for sound to propagate); the structures to accomplish this and other needs[1] developed several times independently in frogs, mammals, reptiles and birds (both sauropsids) [12]. This means findings obtained in different lineages (birds) may not exist in the mammalian brainstem. As we'll see later, this has been a significant complication of sound localization research.

Sound localization also features particular brain circuitry, which has properties not found in other areas of the brain. The need to extract auditory cues (*active encoding*) is another unique feature of sound localization: for sight and somatosensation, the source of a stimulus is always at least partially present in how the sensory organ detected the stimulus in the first place; instead, the tympanic membrane does not explicitly express differences in source position.
Another complication of sound localization research, which has no doubt been one of the driving factors behind its long history, is that the nuclei responsible for this task are situated inside the brainstem, with small, difficult-to-detect potentials, in a location that's difficult to keep in a single brain slice together with its inputs or reach *in-vivo*.

Summarizing, sound localization: (1) is restricted to specific sensory inputs (2) is involved in both main functions of auditory capability (the *where* and the *what*) (3) is implemented by specialized structures (4) needs active encoding (5) is solved by nuclei that are limited in number but difficult to reach and study (6) has practical implications for a growing epidemic of hearing loss. Due to these factors, sound localization is being actively researched for both academia and industry applications (ranging from healthcare to audio products to

[1] We will explore these throughout the text. Among them are comparing arrival times with sub-millisecond differences, analyzing a sound by frequency, and keeping up with sound features using neurons with very high firing rates.

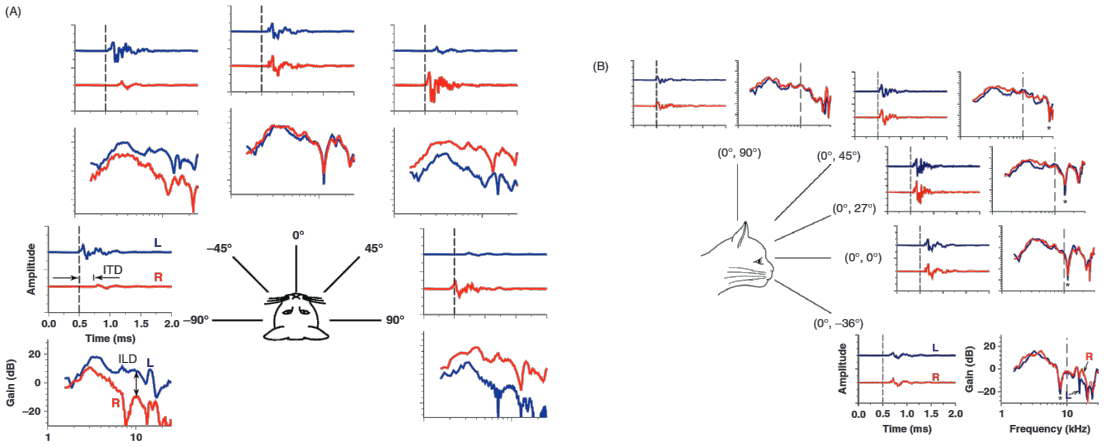


Figure 2: Effect of different azimuthal (A) and elevation (B) angles on sound and its spectrum. The impulse response for how a listener's environment and body composition affect sound arrival times, level, and spectral content is called the Head Related Transfer Function (HRTF). ITD and ILD are visible in (A), while (B) shows the spectral notch (see later). From [68]

videogames[2]). In addition, these features also make it a perfect subject for study using computational methods: its limited focus in terms of input and number of involved neurons (compared to other brain areas) mean that neuronal simulations can be run as a whole; these simulations can be tweaked to experiment with specialized structures, features and theories; the dual interest from industry and research results in many of the components having existing models which strike a good balance between faithfulness and efficiency[3]. In addition, the need for some type of active encoding *justifies* the use of computational models to enable researchers to understand the mechanism behind this encoding. Finally, testing a theory in a simulation before biological experiments limits animal research to a subset of biologically plausible hypotheses.

1.2. Auditory Cues

To derive sound source position, the brain must decode how different positions affect the stimuli received by the two ears. These effects (*cues*) can be classified as binaural and monaural. There are two types of binaural cues: the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD). Monaural (or spectral) cues, instead, are all those experienced by a single ear as a sound moves through space. In this section, we will detail binaural cues by explaining how to calculate them, how they bias evolutionary adaptations, how the brain uses them, and how they can be modeled.

To understand auditory cues, there are two fundamental properties of auditory processing biology that we will mention (detailed explanation in sections 1.4.1 and 1.4.1) before all else: phase locking and tonotopical organization. The first refers to the property of Auditory Nerve Fibers (ANFs) to fire at a specific phase of a sound, hence maintaining information about its frequency and relative arrival phases[4]; this association is maintained until a threshold of 3 to 5 kHz [68]. Tonotopical organization, instead, means that all nuclei involved are organized by frequency; intuitively, higher-level nuclei receive an approximate spectrum of the sound. The two binaural cues we mentioned are part of the "duplex theory", first formalized by Lord Rayleigh [45], according to which localizing sounds on the azimuthal direction is based on these cues. The first cue is the Interaural Time Difference (ITD): the difference in time that the sound wave takes to get to one ear compared to the other[5]. Although simple to understand, it requires special care by the brain, as its full range is only around 1.3 ms (considering an ideal head of width 22 cm); meanwhile, the average duration for a spike event in a neuron is closer to 1 ms[6].
The second binaural cue is the Interaural Level Difference: this represents the shadowing effect that the head has on the level (loudness) of sounds arriving at each ear. The magnitude of this shadowing depends on the ratio

[2]HRTF, defined later in the text, have been used for spatial audio in many videogames.
[3]Some of the peripheral auditory pathway needs to be run real-time on limited hardware, such as cochlear implants
[4]If both pathways are of the same length, then the relative arrival time (or arrival phase) is discernable by comparing the two phase-locked responses. More detail in 1.4.3
[5]Tonotopical organization allows this comparison to happen, as it allows to compare specific frequencies
[6]This means that there must be some mechanism to move from the realm of minute differences in timing ($\simeq 10\,\mu s$) to differences in spike events per second.

between head size and wavelength/frequency: when the wavelength is smaller than the head size, significant shadowing occurs; in general, higher frequencies are shadowed more and lower frequencies are shadowed less 2. Hence, all animal sizes can experience significant ILDs, but the threshold frequency at which these become significant varies depending on head size. This does not mean there is a fixed threshold at which only one of these cues is valid, but their respective effectiveness for each head size rises and falls depending on the frequency.

These binaural cues are not enough to account for all positions in space, but only in the azimuthal dimension: for each pair of ITD and ILD, a cone of possible locations exists that satisfy the ITD and ILD values. In addition, other pieces of psychoacoustic evidence are unexplained by binaural cues, e.g., how sounds heard from speakers are perceived as "external". In contrast, from earphones, they are perceived as "internal". These *missing pieces* are broadly referred to as monaural cues: they account for how the outer ear affects the sound depending on its location. The most common example of the spectral notch: the shape of the pinna introduces a masking effect on a narrow set of frequencies; the frequencies depend on the sound source location, specifically the vertical angle.

The compound effect of delays, attenuations, cancellations, and amplification of a sound traveling from its source to the two ears is called the Head Related Transfer Function (HRTF). Although only the head is mentioned, they account for all the transformations a sound goes through when going from the source to the two ears; as such, HRTFs have three inputs (a source sound, with its azimuth and elevation angles) and two outputs (the sounds as perceived inside the ear canals); depending on the dataset, an HRTF may also need the distance from the listener, while other datasets do not allow for varying distances. Specifically, when distances are above 1 m (far-field region), distance can be accounted for by scaling sound level according to the inverse square law, as HRTFs are asymptotically distance-independent [6]. In contrast, in the near-field proximal region, HRTFs are affected by distance. Measuring an individual's HRTF is time-consuming[7]: it requires multiple measurements, either with in-ear microphones and a speaker or with in-ear speakers and a microphone, with the subject in an anechoic chamber; careful tuning and postprocessing are required to pick a physiologically safe sound level, increase signal-to-noise ratio, suppress background noise and unwanted reflections in measurement results. Once the measurements are complete and fully processed, obtaining the transfer function is easily achieved by moving to the frequency domain. Some generalized HRTF models are available, although they may reduce localization accuracy [66].
The interest in HRTF is not driven only by research purposes but also by industry application: filtering sounds with HRTFs is an efficient way to create virtual acoustic environments (VAEs), which has found its uses in virtual, augmented and mixed reality, but also videogames[8].

1.2.1 Evolutionary perspective and consequences

The relationship between head size and acoustic cue effectiveness has an evolutionary background worth reviewing. To detect and localize air-borne sounds, an animal needs specialized structures (the middle ear) to amplify air pressure waves, and these structures evolved independently in different lineages: mammals, frogs, and sauropsids (reptiles and birds). Fossil studies show that this parallel evolution developed structures (absent in their ancestors) around the triassic [3]. Although this likely means they were under a common evolutionary pressure, it does not mean their solutions for this pressure were the same, as their pre-existing conditions (anatomy and lifestyle) differed:

- Sauropsids were larger, both in overall body size and head size. This meant larger ITD differences, which entails less of a need for highly specialized neuronal processing. It also means a preference for lower frequencies, which makes higher frequencies, where significant ILDs are more likely to occur, less relevant.
- Early sauropsid middle ear (and current frogs) only uses a single middle-ear bone, favoring low-frequency sound conduction. Instead, the mammalian middle ear has always been a three-boned structure favoring higher-frequency conduction [36].

Most likely, the sound localization system in birds (inherited from dinosaurs) evolved to process ITDs to handle low-frequency sounds. Instead, in mammals, where (1) the inner ear does not conduct low frequencies efficiently; (2) ancestors' heads were smaller sizes, which made it much more difficult to detect ITDs; (3) most small mammals today rely on ILDs; mammalian ears likely developed to handle high-frequency sounds, using ILDs, and only later included ITD processing [17].

This evolutionary perspective has two consequences:

Different lineages may have different strategies. There is evidence [10] for a mechanism to detect ITDs in the nucleus laminaris of the barn own. Broadly, using delay lines (i.e., longer and slower axons), internal

[7] Although recently, new methods have risen to tailor HRTFs to individuals in home environments [66], more in [66]
[8] In Valorant, Riot Games even calls its setting "HRTF" [42]

delays are created, which compensate the external ITD; then, the targets can act as coincidence detectors (a known feature of neurons), where only the one with the correct (i.e., corresponding) delay fires, hence identifying the current ITD. This type of processing, the corresponding map of internal delays, and its related properties were not found in mammals [68]. Still, mammals recognize ITDs, so they must use a different strategy to perform the same function.

Different nuclei are dedicated to different auditory cues. As we will see in the next section, ILDs and ITDs are primarily processed in different nuclei with a shared pathway. Specifically, the structure processing ILDs, where binaural inputs first meet, is the lateral superior olive (LSO), and it is homogenous in all mammals investigated [59].

1.3. Input types and behavioral results

In this brief section, we introduce sound types, which we will reference throughout the text, and recap some of the behavioral results that will inform the evaluation of our results. For our concerns, sounds can be classified by frequency and level. Humans can hear sounds ranging from 20 Hz to 20 kHz, with higher frequency hearing degrading most with age. Safe levels for humans are up to 90-100 dB. Pure tones are sounds with a sinusoidal waveform, so they have constant frequency, phase shift, and amplitude. Because of these features, tonnes will be featured most prominently in our work since they allow us to selectively investigate a specific frequency due to the tonotopical nature of sound processing. The two other sound types featured in this work are white noise and clicks. White noise is a random signal with equal power at all frequencies, making it perfect for masking other signals. Lastly, clicks are brief, broadband (thus stimulating the entire frequency spectrum) stimuli, sometimes called transients, which can be used to study timing.
In behavioral studies, humans show higher accuracy in the localization of broadband signals than pure tones and cannot achieve vertical localization without spectral cues. As expected from section 1.2, front-back confusions are common [22] and sometimes counted as correct in early results. It is unclear whether accuracy improves with louder sounds (within safe thresholds) or longer expositions [69]. The sound localization ability is also studied regarding its sensitivity, i.e., the just noticeable difference (JND), usually expressed as the minimum audible angle (MAA), from a reference sound. In a critical result by Mills, the MAA was measured as a function of the azimuthal position of the reference sound: the MAA systematically grew as the reference sound moved away from zero degrees (in front of the subject).

1.4. Sound Processing

This section will review the auditory pathway and how to model it. It will be organized in parallel, pointing out relevant anatomical features and explaining how they can be modelled. Because our focus will be on the features relevant to sound localization, much will be omitted. At a high level, the auditory pathway begins outside the body, as the environment and the outer ear modify sound, which causes the eardrum to vibrate. This vibration is analyzed, modified, and transmitted as a graded potential, which becomes the input for the first neural component. Further neural connections then enhance the signal and transmit it to the processing nuclei in the brainstem. We split the pathway into three sections: first, the peripheral system, which is composed of all sound processing structures that are not neural in nature; then, the monaural neural pathways, which filter and convey sound signals to the processing nuclei; finally, the binaural section, where inputs from the two ears first meet, which perform part of the comparisons and terminate in the brainstem.

1.4.1 Peripheral System

The sequence of transformations starts outside the body, as sound is affected by the head, torso, and outer ear until pressure waves cause the eardrum to vibrate. Small ossicles in the middle ear amplify and transmit the eardrum vibration to the cochlea. The vibration becomes fluid waves in the cochlear fluid, which causes the basilar membrane (BM) inside the cochlea to respond; the movement of the BM is further modified by outer hair cells (OHCs) and detected by inner hair cells (IHCs), which transduce the motion into fluctuations of a graded potential which controls neurotransmitter release. These neurotransmitters activate the first neural step, the auditory nerve fibers (ANFs).
In the following paragraphs, a commonality of all modeling work arises: modelers must strike a balance between white-box, bioplausible (mechanical or otherwise), computationally expensive simulations and grey-box system identification to obtain computationally less expensive, more opaque models.

Outside and Outer ear We have already mentioned how traveling from sound source to the two ears changes the sound: due to diffractions, scattering, interference and resonance, by the body and obstacles,

attenuation, cancellation and amplification can be observed, and the spectral content of the signal changes. Two possible routes have been explored to reproduce HRTFs: simulating a physical model or creating a digital filter. By simulating a physical model of the sound waves interacting with anatomical features and the room [64], it is possible to not only understand how each feature affects the result but also to adapt it to different subjects or environments; these methods are very computationally expensive. On the other side, it is possible to create a digital filter that reproduces the sampled experimental data [31], where the complexity is always lower and can be further lowered with simpler filters (provided they capture sufficient features of the HRTF). Although vastly more efficient, these digital filters do not provide any insight into the effect of individual anatomical components.

Middle Ear As the eardrum vibrates, three ossicles, the malleus, the incus, and the stapes, transmit the vibration to the oval window. A combination of the low surface area ratio between the eardrum and stapes footplate and the lever ratio of the ossicles allows the middle ear to move the low-impedance (air vs. fluid), low-pressure, large-displacement vibrations into higher-impedance, higher-pressure, small-displacement vibrations in the cochlear fluid[9]. The middle ear does not introduce distortions at physiologically safe levels, but the filtering effect is not linear [49]. Modeling approaches follow under three categories: lumped-element models [47], biomechanical models [15], and digital filters [58][23]. The first use the parallel between acoustic and electrical elements, and consider the middle ear a transmission line with lumped electrical models of elements; this was the classical modeling technique, and it is now less common. Biomechanical models reconstruct the geometry of the elements in a physics simulation and (as in the outer ear) are computationally expensive but allow for fine-grained investigation of healthy or impaired tissues on sound transmission. The third class of models, digital filters, use either filter cascades or a single filter to achieve realistic frequency responses, multiplied by a scalar for a realistic gain [40]. Some authors [11][58] suggest that the middle ear is responsible for the asymmetries in basilar membrane responses (both for iso-intensity curves and so-called glide or chirp, more on this later).

Basilar Membrane and Outer Hair Cells As the stapes vibrates on the oval window, its vibrations are transmitted to the cochlear fluid, creating a pressure wave. This fluid is contained in two connected chambers (through the helicotrema) at the apex, the scala vestibuli and the scala tympani. The endings of this two-chamber system are both sealed but flexible: on the side of the stapes, the oval window can flex (to transmit stapes vibrations), while on the opposite side, the round window is flexible as well. This allows the pressure wave to be created. As the fluid (incompressible), moved by the wave, pushes and pulls on the bottom of the chamber (basilar membrane), the basilar membrane moves up and down, creating what's called the "traveling wave". Because of varying width and stiffness, different points of the BM respond to different frequencies: the base of the BM responds to high frequencies, while the apex vibrates at low frequencies, with frequencies arranged logarithmically [26]. Thanks to this, the inner ear can organize sounds by frequency (performing an approximate spectrum); this distribution is called tonotopic organization.[10]. The BM is also the base of hair cells, and each placement of hair cells is usually called "site". There are two types of hair cells: inner hair cells (IHCs) and outer hair cells (OHCs). The first (IHCs) are responsible for transducing mechanical signals into electrical; the second (OHCs) are the active component of the cochlea [24], as they modulate basilar membrane motion, both increasing sensitivity and enabling frequency selectivity in complex sound environments [48]. We will consider this active effect part of the BM modeling, as this is the most common approach.
Each BM site responds to a small range of frequencies, centered around the frequency of maximal response, called *characteristic frequency*(CF), with responses tapering off at nearby frequencies. Due to this, every BM site is modeled as a frequency filter, with the whole BM being a series (*bank*) of overlapping filters. These filters are asymmetric in two dimensions: (1) the response strength of a single site increases slowly for frequencies below the CF but tapers off quickly above it; (2) the instantaneous CF of a single site changes during a brief stimulus (like a click), increasing at the basal site and possibly decreasing at the peak ([8] vs [32]), a phenomenon called chirp or glide. There are various nonlinearities in BM filter response [35]: they show more gain at low levels than high (growing compressively with level); both CF and bandwidth (how wide the set of nearby frequencies the site responds to) change depending on level; two-tone suppression (the response to one tone decreases when paired with a different tone); distortion. These responses depend on the physiological state of the cochlea: as they depend on active components (OHCs) [24], they become fully linear post-mortem.
BM model classes are the same as previous components (transmission line, biomechanical, digital filters). Given the breadth of approaches, we will focus on the ones relevant to our investigation: digital filters. The most common classical model is the gammatone filter 3, developed to simulate auditory nerve fiber responses. The response of the gammatone filter is the product of a tone (at frequency equal to the CF of the BM site) and

[9]Otherwise, sound airwaves would just bounce away: try screaming at a friend underwater.

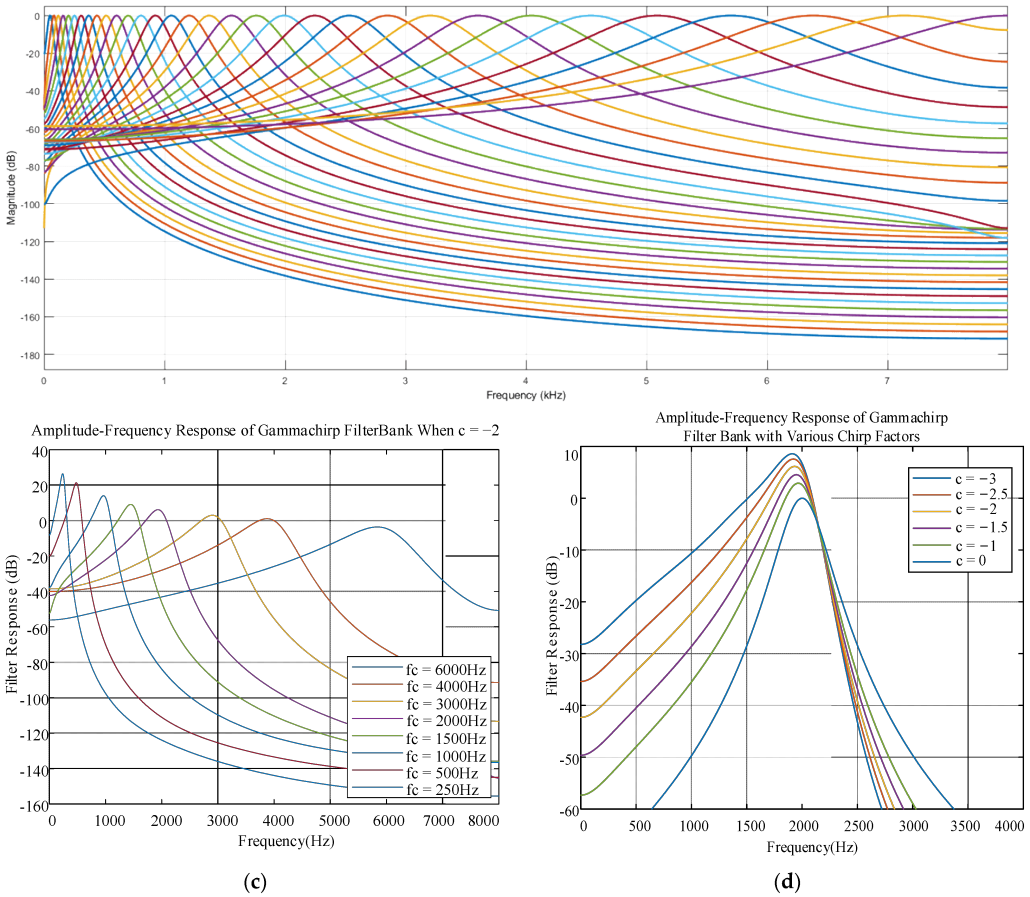
[10]The traveling wave starts at the base of the BM, increasing in amplitude until it gets to the frequency of the input sound

a gamma distribution that determines how adjacent frequencies are also involved. To model the entire BM, it is approximated to a series of gammatone filters with increasing CF, a filter bank. This filter is linear, and symmetric in its response. Improvements to the original gammatone led to the gammachirp filter, which was not symmetric in its response, by modulating the aforementioned tone in frequency: from here, the chirp name. After further improvements, the filter became the compressive gammachirp [25], composed of three simple filters: first, the gammatone filter followed by a low-pass filter, which combined produce the asymmetric gammatone-like filter; then, a high-pass filter modulated by frequency, which accomplishes the level-dependent gain (compression).
The compressive gammachirp accurately reproduces human auditory filters over many center frequencies and levels, and "could probably be used to simulate physiological BM iso-intensity responses directly, although no studies have been reported to date aimed at testing the filter in this regard" [40]. Although the frequency glide is included, the trends (the direction of the shift compared to level) are inconsistent with physiological data. It is likely unable [44] to reproduce two-tone suppression and more advanced combination effects. More advanced models exist: the Tan-Carney model [58], a gammatone filter where gain and bandwidth vary dynamically to account for the nonlinear properties; the Dual Resonance NonLinear (DRNL) [34], composed of the sum of a linear and a nonlinear pathway, which can also be adjusted to simulate OHC loss (more detail later). Because of our narrow focus on sound localization, and since our investigation will primarily focus on single tones, the most complex and faithful models are outside the scope of this work: the features they reproduce are not shown in higher localization neural nuclei, and their computation time is much higher than simple models.

Inner Hair Cells As the cochlear fluid and the BM of a single site vibrate, the hair cells in the organ of Corti are moved back and forth. Specifically, the inner hair cells (IHCs) are responsible for the transduction of mechanical to electrical signal. IHC stereocilia are organized in bundles, with their tips connected by links: when deflected towards the tallest of the bundle, the links pull open ion channels (potassium) to increase the inward flow of ions that depolarize the cell; when deflected in the opposite directions, ion channels are closed to prevent the flow of ions inside the cell. As the voltage inside the cell changes, voltage-gated calcium channels are opened and closed, and calcium ions promote the release of neurotransmitters (glutamate) into the synaptic cleft. This *graded potential* is then picked up by the auditory nerve fibers (ANFs). This means the signal is only transduced in one direction (asymmetric gating of the ion channels), which gives the IHCs the characteristic feature of half-wave rectifiers. This also means that the concentration of ions will be at maximum at the moment corresponding to a specific sound phase, which causes the following neurons to be *phase locked*. This results in a complex environment where the intracellular voltage is determined by (1) inward potassium flow, driven by stereocilia movement (2) outward potassium flow, to compensate the inward, at the basolateral membrane (3) inward calcium ions, that cause the neurotransmitter release (4) a capacitive effect of the IHC membrane (which results in a low-pass filter)[11] (5) the homeostasis of the organ of Corti. In addition, it is difficult to isolate the contribution of the IHC from those of the middle ear and the BM; this contribution is estimated by measuring the growth of the AC and DC components while changing the level of an input sound at a frequency **below** the CF so that the BM effect is almost linear (the middle ear, as we said, stays linear). The result also reflects the effect of the IHC membrane (and likely other factors [65]), which limits the AC effect (phase locking) to frequencies below 3 to 5 kHz; below this threshold, the AC component matches the stimulus frequency, while the DC component matches the stimulus level. As can be imagined from our description, a biophysical approach is to model the IHC with an electrical circuit: this results in accurate response characteristics across different datasets [33]. Alternatively, digital filter approaches use a series of asymmetric gains (up to a saturation point), followed by a low-pass filter for the capacitive effect of the membrane. Their shortcomings are as earlier: they do not allow for physiological inspection, cannot model non-healthy IHCs, and are less accurate than biophysical models.

Auditory Nerve Synapse The connection to the auditory nerve is achieved by releasing glutamate (the neurotransmitter) into the synaptic cleft. This depends both on the voltage of the IHC and the availability of vesicles in the presynaptic area. During an extended stimulus, lower availability will cause lower activation, which causes the AN spike rate to fall. This phenomenon is called "adaptation". Hypothesizing a single reservoir of vesicles would entail a single time constant of adaptation; instead, due to data indicating multiple constants, two similar models have been proposed: the first based on multiple successive reservoirs [9], the second on reuptake from the synaptic cleft [39], and as reviewed by one of the authors the models are mathematically similar [70]. The exact connection between vesicle release and AN spike is unknown [40], so it is usually modeled probabilistically. In our case, this probability will be collapsed in a realization, as we need real spikes for the remainder of the network.

[11] Consider a simple half-wave rectifier with a capacitance: as the frequency rises, between cycles the capacitance will have to provide current for less and less time. Hence, high frequencies will be masked, while low frequencies, in which the capacitance has time to deplete, will be reflected in the output.



(c) (d)

Figure 3: Differences between gammatone (top) and gammachirp (bottom) amplitude/frequency responses. As the characteristic (central) frequency increases, the bandwidth of the filters increases. CFs are spaced exponentially, but in gammatone filters the frequency response of each filter is symmetric about the CF. Gammatone filters also do not modulate peak response. From MATLAB documentation [1] and [20].

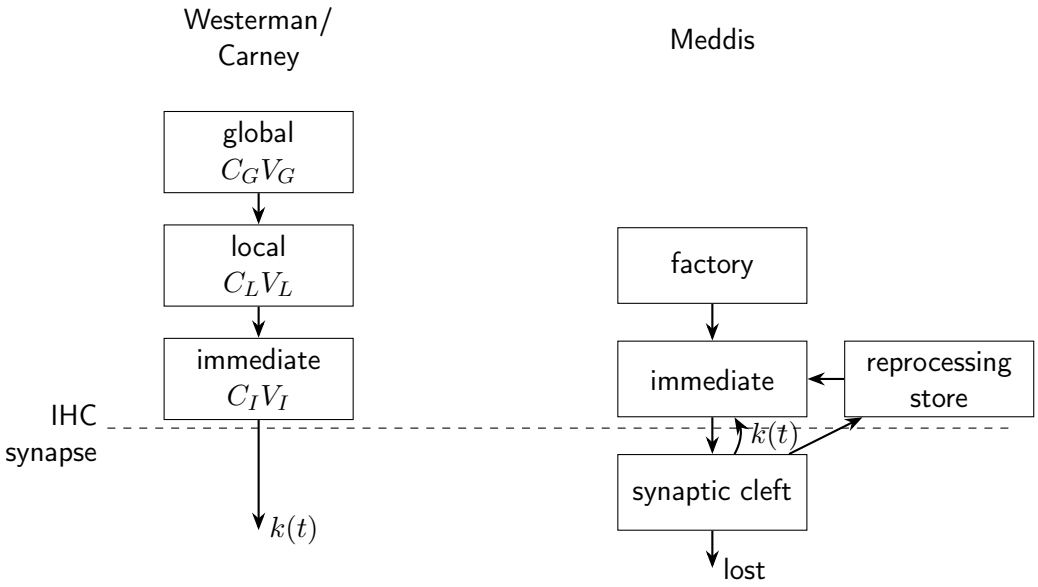


Figure 4: The Carney model for IHC-ANF synapse compared to the Meddis model. From [40]

1.4.2 Monaural Neural Pathway

Starting with the ANFs, the input becomes neural in nature. These inputs are filtered by bushy cells, then faithfully carried to both sides to be processed by higher centers in the superior olivary center (SOC). We will focus on the pathway generally associated with azimuthal localization, leaving aside the pathway handling spectral cues for the vertical direction. Monaural signals are then integrated by the lateral superior olive (LSO) and the medial superior olive (MSO), respectively associated with ILD and ITD.

Auditory Nerve Fibers So far, we've only considered afferent ANFs (from the cochlea to the central nervous system), but about 10% of ANFs are efferent going from the brainstem back to the cochlea. The mechanism driving efferent fibers is still unclear and rarely included in models, so they are outside the scope of this work. Afferent fibers are classified in type I and type II, respectively 90% (innervating IHCs) and 10% of ANFs (for OHCs). Very little about type II fibers is known, but recent evidence suggests they may respond in case of OHC damage, carrying nociceptive information. Our focus will be on afferent, type I ANFs.
Every IHC is innervated by ten to thirty type I ANFs, summing up to about 30,000 in total [55]. ANFs innervate a single IHC, maintaining the IHC frequency tuning and showing phase-locking, and generate spontaneous spikerates (SR) of up to 100 spikes/s. All ANFs terminate in the cochlear nuclei (CN): after entering through the nerve root area, they bifurcate, with one branch going rostrally[12] to the antero-ventral cochlear nucleus (AVCN) and the other caudally[13] to the postero-ventral cochlear nucleus (PVCN) and then the dorsal cochlear nucleus (DCN). The latter's functions are coding for spectral cues for vertical sound localization, filtering uninformative self-generated sounds, and processing of ultrasonic vocalizations [68]. Although there are multiple cell types in the AVCN, the majority (and what we will focus on) are bushy cells, named after their bush-like dendritic tree. There are two types of bushy cells: globular bushy cells (GBCs) and spherical bushy cells (SBCs).

Bushy cells Due to a specialized set of currents, bushy cells have a very short membrane time constant, hence the quick decay of EPSPs. The fast decay entails that for bushy cells to spike they need either many coincident inputs or few very large inputs. Other cells, such as MSO, LSO, and cells in the MNTB, have since been discovered to have similar properties [30][68]. SBCs receive one to four of the largest ANF terminals (end bulbs of Held) on their soma, with large, often suprathreshold EPSPs generated. Due to this, their response is very similar to ANFs, a response called primary-like [43]; this also means that SBCs stay phase-locked at low frequencies, remarkably even more accurately than ANFs [27]. Axons of SBCs then project directly to the ipsilateral LSO and both MSOs. GBCs receive more ANF terminals (20/cell or more), but they are smaller and usually generate subthreshold EPSPs on their own. Hence, GBCs spike when multiple inputs arrive simultaneously, producing a primary-like-with-notch response: at the onset of high-rate input, they fire an initial spike and then pause due to the refractory period before spiking again. At low frequencies, they remain phase-locked. GBCs form synapses with glycinergic (inhibitory) cell bodies in both sides of the brain:

[12] In the general direction of the face
[13] In the general direction of the back of the head

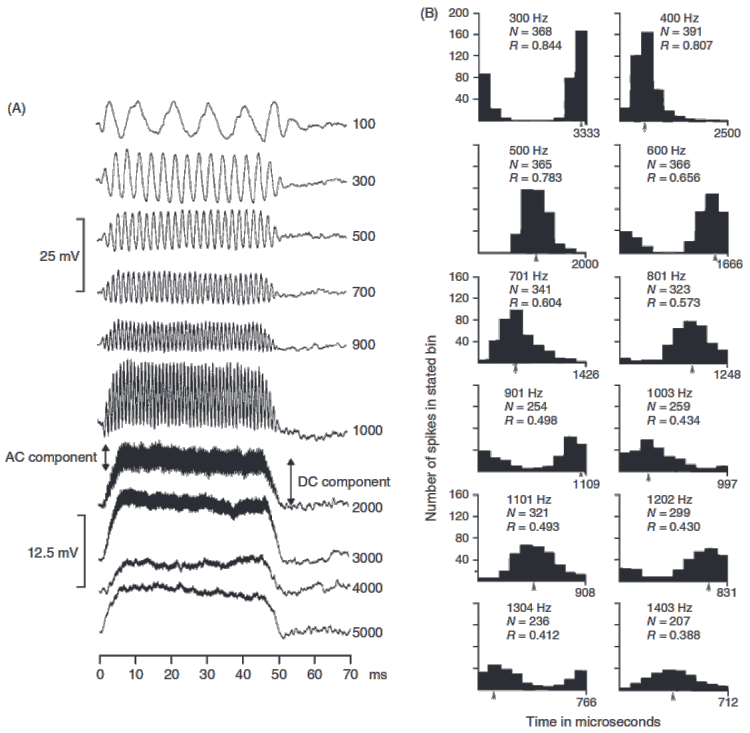


Figure 5: How phase locking and vector strength evolve with changing frequency/CF ratio in IHC (left) and ANF (right). In (A), the intracellular potential of an IHC from a guinea pig presented with tones at 80dB, varying in frequency. In (B), spikes from a squirrel monkey stimulated with tones organized in histograms show how much the ANF was phase-locked to a specific phase. Vector strength is noted in each. The CF of the fiber is 1100 Hz. From [68]

ipsilaterally, they connect to cells in the lateral nucleus of the trapezoid body (LNTB); contralaterally, they form the largest synapses of the central nervous system on cells of the medial nucleus of the trapezoid body (MNTB).

MNTB and LNTB. As bushy cell axons travel towards the brainstem, they reach the superior olivary complex (SOC). The SOC contains multiple nuclei that have roles in hearing, both in ascending and descending pathways. The LNTB, MNTB, LSO, and MSO are the most well-defined and studied. The MNTB features the largest synapse in the brain, the calyx of Held, so large that recording from both of its sides is possible. Both MNTB and LNTB provide inhibition to their targets: ipsilateral LSO and MSO for the MNTB and ipsilateral MSO for the LNTB. The MNTB shows similar responses to their GBC inputs, with short membrane time constants, and the LTNB shows a variety of primarily monaural responses [19][61]. While some LNTB neurons show sensitivity to ITDs [14], the mechanism behind it is unclear (as is whether the sensibility is generated or inherited).

1.4.3 Binaural Cues Processing

The remaining nuclei of interest, the LSO and the MSO, are where bilateral inputs finally meet. LSO cells are excited by stimulation of the ipsilateral ear and inhibited by the contralateral ear; MSOs respond to stimulation of both ears. This section will overview their physiological properties and discuss what they mean for sound localization.

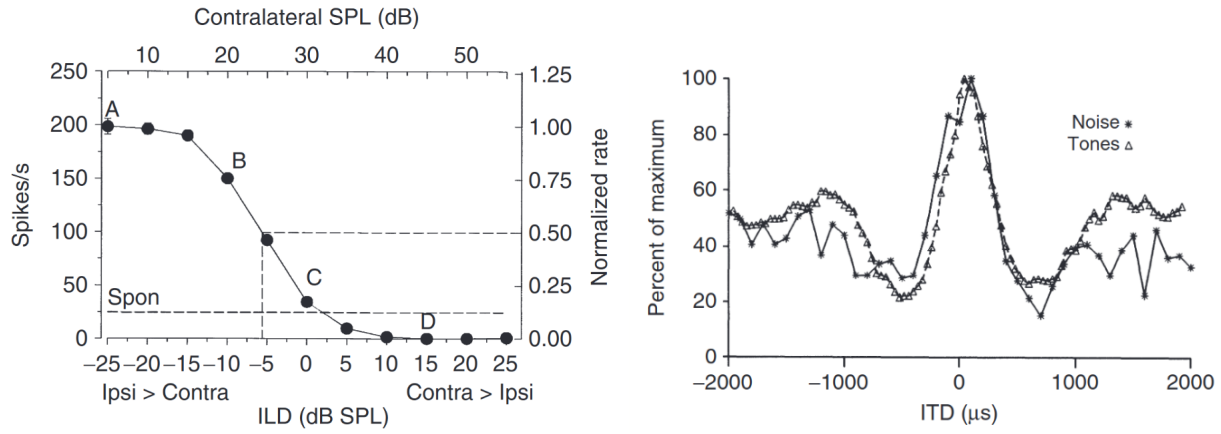


Figure 6: A plot showing the response of SOC nuclei to the cue they are most sensitive to, adapted from [68]. On the left, the response of an LSO cell to an increasing contralateral level (increasing ILD). On the right, the response of an MSO cell to varying ITD. Notice the marked ipsilateral preference for the LSO, and contralateral-favoring response of the MSO.

LSO Although there are four total inputs to the LSO, the two strongest inputs are the ones involved in sound localization: ipsilateral SBCs make excitatory synapses, while neurons from the MNTB form inhibitory synapses, which carry information from the contralateral GBCs. Most LSO axons then project to ipsi or contralateral ICC and DNLL; neurons projecting to the ICC innervate regions overlapping with those innervated by MSO neurons.

MSO The MSO's principal cells have a strongly bipolar morphology, with dendrites stretched horizontally receiving bilateral inputs, both excitatory and inhibitory. The primary excitatory inputs to the MSO are from the SBCs of both sides, segregated on the two primary dendrites; instead, the primary inhibitory inputs come from the MNTB and LNTB of the side of the LSO, and synapse on the soma. The inhibition from the MNTB carries the signal from the contralateral ear. As mentioned, the MSO then projects to, among other nuclei, the IC, where the synaptic terminals synapse in regions overlapping *ipsilateral* LSO terminals. Because of the small amplitude of responses, MSO recordings in vivo are difficult to obtain. The MSO presents a low input resistance and fast membrane time constant, requiring precise coincidence of subthreshold EPSPs for spiking. IPSPs have slower kinetics than excitatory inputs. An important characteristic is that stimulating ANFs, both contralaterally and ipsilaterally, causes an IPSP which arrives at the MSO before than the EPSP: this is suprising, because the excitatory path has a single synapse inbetween (ANF-SBC), while the inhibitory path has two synapses (ANF-GBC-MNTB).

As we overviewed in 1.2, two main cues are involved in sound localization: interaural level disparities and interaural time disparities. The two nuclei we mentioned are the most likely candidates for where the processing of these cues happens.

Processing of ILDs Computing the ILD means subtracting the magnitude of the response of one ear from the other. As we mentioned, the LSO is excited by stimulation of the ipsilateral ear (from the SBCs) and inhibited by the contralateral ear (from the MNTB); this is very likely to be the main mechanism for processing ILDs. In addition, ILDs were shown to be the cue that LSO cells are most responsive to [60]: as said in 1.2, recording responses to free-field inputs is complicated, so most recordings are obtained using virtual acoustic spaces (VAS), partial HRTFs which only include some cues. The LSO is still tonotopically organized [62], so it is important to consider that the ILD the LSO can compute will still be dependent on frequency: a single location of the LSO receives intensity information of a small range of frequencies. The size of the range depends on (1) how sensitive a specific BM region is to frequencies different from its CF (modeled with the gammatone bandwidth) (2) how many BM regions are sensitive to each frequency (modeled with the density of gammatones and the ERB scale, see later) (3) how many ANFs converge to a single SBC, so activity in different BM regions collapses to the same SBC (convergence). In the LSO, the latencies of inhibition and excitation, respectively contralateral and ipsilateral, are similar, even though inhibition requires an additional synapse through the MNTB.

Processing of ITDs Unlike the processing of ILD, the processing of ITDs is less understood. This may be in part because the nucleus that is believed to process them is difficult to record from: it is narrow, sometimes

one or two cells thick; its action potentials are small; the region is flooded with phase-locked potentials from the auditory chain we've reviewed so far [51]. Multiple strategies have been proposed, but the precursor 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 (see 7).

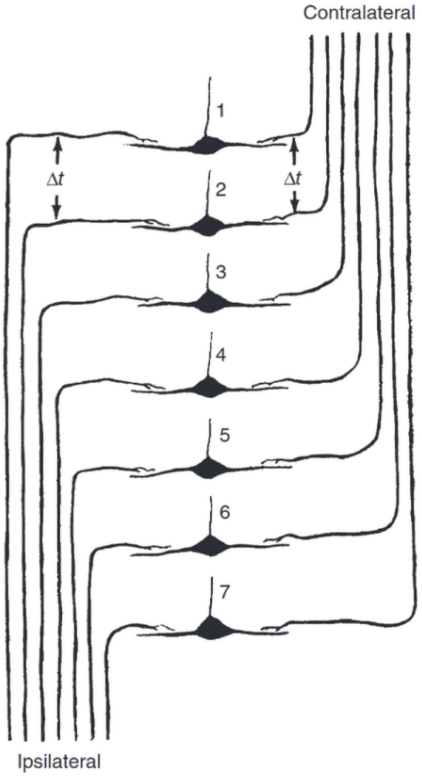


Figure 7: The Jeffress model to recognize ITDs, from [68]

.

If we consider the MSO to be the site of this model, its first assumption is that precise timing information is available to the MSO: this assumption is correct, as the mechanism of phase-locking can, within the frequency range in which it functions, relay precise information about the phase (and hence the timing) of both sides. In addition, SBCs and GBCs enhance ANFs phase-locking properties [28]. The role of the coincidence detector that the Jeffress model would assign to MSO is also reasonably accurate: it responds maximally when inputs arrive almost coincidentally. This difference from the exact coincidence varies among cells and makes for a cyclical response pattern modulated by frequency. ITD curves for the same neuron at various frequencies share a characteristic delay (CD) at which they all have the same relative amplitude. If the conduction time were constant with respect to the frequency of the input, CDs would only be at peaks or troughs, but this is not the case: while most cells in the MSO have CDs when all curves are at their peak (thus called peak-type cells, figure A), others have their CDs at different phases (non-peak type).

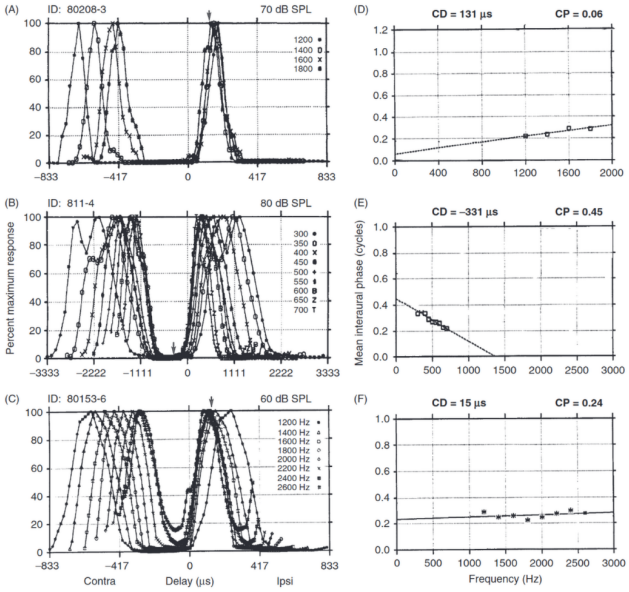


Figure 8: Responses from the ICC, used as a proxy for the MSO, showing the different types of cells. On the right, the interaural phase vs frequency plots. ITD curves are normalized. From [68]

This is where the Jeffress model shows its weaknesses: (1) only peak-type cells make sense in a Jeffress interpretation (2) the model would predict neurons to show a range of internal delays, distributed on the range of possible ITDs (3) for the cause to be the axonal delay due to its length, these delays would need to be distributed spatially. Consequence (2) would expect delays to correspond to the physiological ITD range. Instead, evidence in cat [67] and guinea pig [37] IC (used as a proxy for the MSO) show a similar range despite much different head sizes. In addition, although axonal length varied, a re-analysis of past data [29] found two important inconsistencies with consequence (3): the anatomical data could not explain the range of optimal delays (OD, maximal response ITD), and there was an inverse relationship with CF (large ODs only in low frequencies, and OD decreased as frequency increased [37][14]) which would not be compatible with a delay line mechanism (but more compatible with an internal phase-shifting mechanism). We are left with two important questions, to which there have been multiple plausible answers but no definitive ones:

- **How are internal delays generated?** There are two main lines of research; (1) *Cochlear delays*: connecting locations in the basilar membrane that resonate at different frequencies, one could make use of the relatively slow traveling wave to generate delays in the cochlea itself [68]; this theory has very limited physiological tests, and to correspond to physiological data, where CDs are biased toward the contralateral ear, errors would need to be biased as well, with the contralateral CF being lower than the ipsilateral CF. (2) *Inhibition-driven*: as we've mentioned, several factors suggest an essential role for inhibition; among them: the bipolar nature of MSO neurons; two different sources of inhibition, one from each ear; inhibition is phase locked; the synapse driving the contralateral inhibition (the calyx of Held, from GBCs to MNTB) is the largest known synapse; inhibition was shown to be faster (in arrival time) than the corresponding excitation; blocking inhibition shifts ODs towards 0ms [5]. Yet, the mechanism behind the role of inhibition is unknown; one proposal was that MNTB-carried inhibition delays the excitatory EPSP [5]; in one later study, inhibition was found not to have any effect on ITD tuning [46], while in another one preceding inhibition could not be measured [63]. A following proposal [41] explored the effect of various time deltas between each pair of excitation and inhibition from the same side and found that the ITD of maximal response could be shifted significantly by changing these two parameters, highlighting the relevancy of both inhibition sources.
- **What is their relationship with ITDs?** The relationship between internal delays, optimal delays, and species ITD is still unclear. As mentioned, species with different head widths (hence different maximal ITD) do not show significant differences in the distribution of ODs; in addition, OD distribution is not restricted to the physiological range of the species. This has led researchers to consider the possibility that ODs are not distributed to respond maximally to specific ITDs but to have the maximal change in firing rate over the species physiological range for ITDs [38][21]. This led to the *two-channel model*, according to which the activity of LSO and MSO can be combined, with opposite maximal response points (9). The original formulation of the two-channel model (encompassing the integration of LSO and

[14]While the trend was confirmed in other animals, the lack of small ODs in low frequencies wasn't.

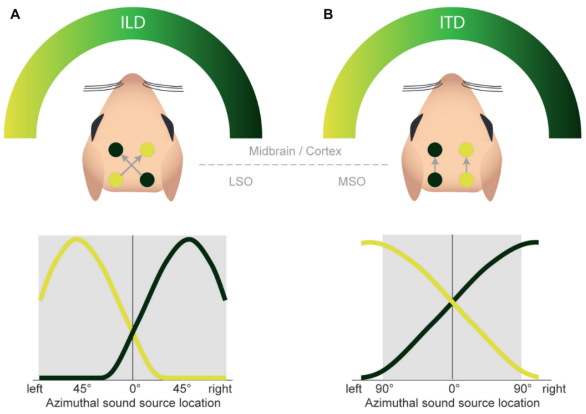


Figure 9: A diagram showing how LSO and MSO inputs converge to higher centers, in the two-channel model. From [18]

MSO) is incompatible with several results[15], with some authors [17] now requalifying the MSO as focused on sound segregation, due to its context-dependant modulation.

Inferior Colliculus Finally, we examine some properties of the last site of convergence we will cover, 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 (BIN). The ICC receives input from both of the SON 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) [18]. 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. Some results [53] suggest that most cue integration happens at the BIN, but the same authors found that most neurons in the BIN responded strongly to ILDs, while the contribution of ITDs was weak [54]. In addition, evidence for a spatial map has not been found, and some results [52] 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.

[15] See [68] for details; as examples: unilateral midbrain lesions leave most subjects with unaffected ipsilateral localization ability but impaired in contralateral space; the variety in MSO ODs, considered noise for a two-channel model, and the closeness to zero ITD are important features for interaural decorrelation, very acute in humans

2. Aim of the Thesis

Given the state of research around sound localization, this work improves an existing, simplified, neural-only computational model [50], with three main objectives:

1. Improve the bioplausibility of the existing model by:
   (a) Implementing peripheral sections at three levels of bioplausibility: (1) *Pulse packet* (see 3.1.4) (2) *Gammatone-based* (3) *Tan-Carney based, bioplausible mammalian cochlea*
   (b) Evaluating the effect of these peripheral sections on SOC nuclei
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*: JND 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 into two sections: the first focused on increasing the bioplausibility of the peripheral system, and the second increased the bioplausibility of the neural processing. As such, the needs of the two sections varied substantively, and different strategies were adopted. The modeling of the peripheral section, from the sound to the spiking output of ANFs, was implemented using Brian2Hears, an extension to Brian 2. The modeling of the central nervous system was developed using the NEST simulator framework. This section will introduce the two, explain the differences, detail how the simulation was set up, and define how we will test its results. We will take a sequential approach to mirror the biological explanation, starting with the peripheral simulation and then explaining the central processing.

Considering what has been mentioned so far, we will use these conventions:

- Only consider front-facing angles, where zero degrees represents the center, -90 is extreme left, and +90 is extreme right.
- Only consider human application: all ITDs shown will use a human range, and, where this is not specified, the diameter of the human head will be approximated as 22 cm

3.1. Peripheral Modeling

3.1.1 Simulator

We modeled the bioplausible peripheral stages using Brian 2 Hears [13]. Brian 2 Hears is an extension for Brian 2 [57] to enable auditory modeling; it is also usable as standalone, and we will use it both ways. Brian 2 Hears and Brian 2 are both written in Python. As we will see in section WRITEME, one important weakness of Brian 2 Hears is that it is inherently single-threaded: although simulations can be run in parallel, a single simulation cannot be sped up by parallel computation.
Using Brian 2 Hears, we implemented two peripheral sections at different levels of bioplausibility. As we mentioned in section 1.4, components of the peripheral auditory system can be modeled using different strategies; using Brian 2 Hears means utilizing digital filters to approximate the behavior of each component. We also detailed how modelers must evaluate the complexity/bioplausibility tradeoff for each component: this would have resulted in far too many models to test end evaluate, so we decided to limit the choice to a simple gammatone model and the more advanced Tan-Carney model [58]. We also avoided optimizing model parameters, and used literature values.
Figure 10 shows the pathway common to all peripheral models. First, a sound, together with its location, is filtered through an HRTF, obtaining the filtered sounds as they arrive in the acoustic canal of the two ears. Then, each ear goes through peripheral processing, which varies among the different models we tested, and causes an ANF spiking pattern. The ANF spike trains are the final result of the peripheral processing pipeline.

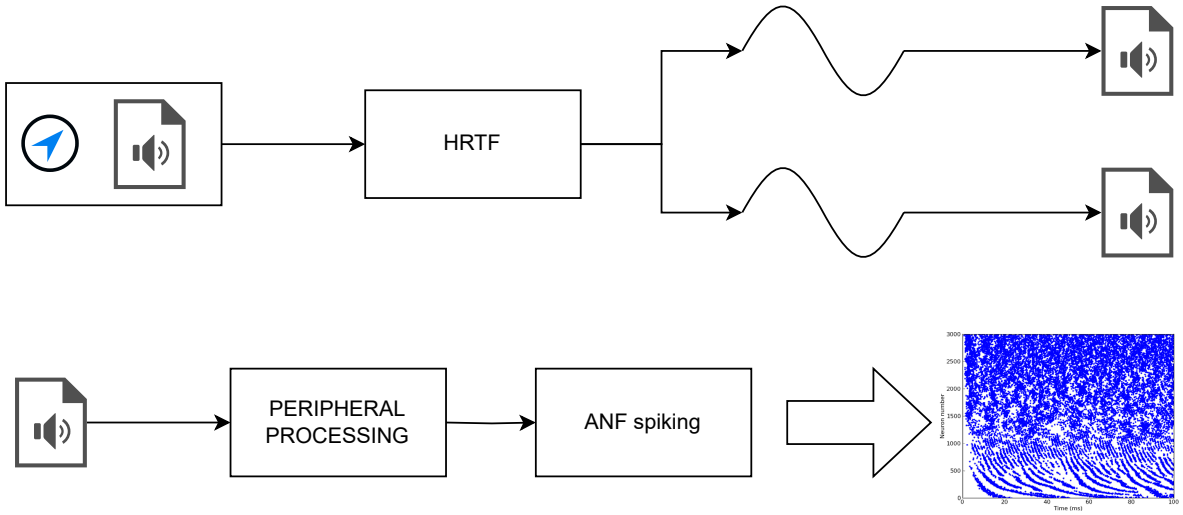


Figure 10: Block scheme for auditory peripheral processing

3.1.2 Gammatone

The processing of the Gammatone peripheral model is articulated in these steps:

1. *HRTF*: the sound is filtered through an HRTF to obtain the two inputs to each ear. We used the IRCAM LISTEN [2] database. This early database is well-integrated with brian2hears[16]. This limits us to 15° increments in azimuthal position (13 positions). This process is common to all our peripheral models.
2. *CF spacing*: as we mentioned, CFs lower logarithmically along the BM. This organization is approximated using the center frequencies of the ERB scale [72]. We used 3500 IHCs, as in the human cochlea [4].
3. *Gammatone*: then, for each ear, CFs become the center frequencies of 3500 gammatone filters in a filter bank. The sound is filtered by the filter bank in each ear, modeling how inner hair cells are engaged depending on frequency.
4. *Cochlea*: After the gammatone filterbank, the active component of the cochlea is modeled with a cube root power law, the simplest estimate [56] of the active cochlea compression.
5. *Current transformation*: Each filtered response will now be interpreted as a current raising the cell voltage of an IHC cell. To do that, it will be rectified (half wave). This will be analogous to the mechanical movement of the IHC opening and closing ion channels. The current is then multiplied by an appropriate scalar (15).
6. *ANFs*: The current then raises the voltage of IHC cells, causing it to release more neurotransmitters in the synaptic cleft, making the ANFs spike. Effectively, the IHC and ANF are modeled as a single unit whose inner voltage is raised by the current until the spike threshold is reached. This is where noise is included in the output, summed to a simple leaky integrate-and-fire model. This neuron model includes a refractory period: when the refractory period is set at 1 ms, it will not be able to fire faster than 1 kHz. This can be used to model effective phase locking at low frequencies and the inability to cope with high frequencies.

This results in a model which shows phase locking (particularly obvious at low frequencies), but will have limited biological plausibility, showing no adaptation (as it has no concept of vesicle availability), no tone interaction, approximate IHC frequency response and the other drawbacks explained in section 1.4. At the same time, most of these features are not strictly related to sound localization. As this model is not attributed to a specific author, we experimented with noise level. This noise models the spontaneous rate, not the noise in the sound. All effects on the sound are carried by the HRTF as if our experiment *in silico* were happening in an anechoic chamber.

[16]Its recording setup was a single-microphone, single-speaker setup, with the speak attached to a metallic u-shaped supporting structure with two dimensions of movement.

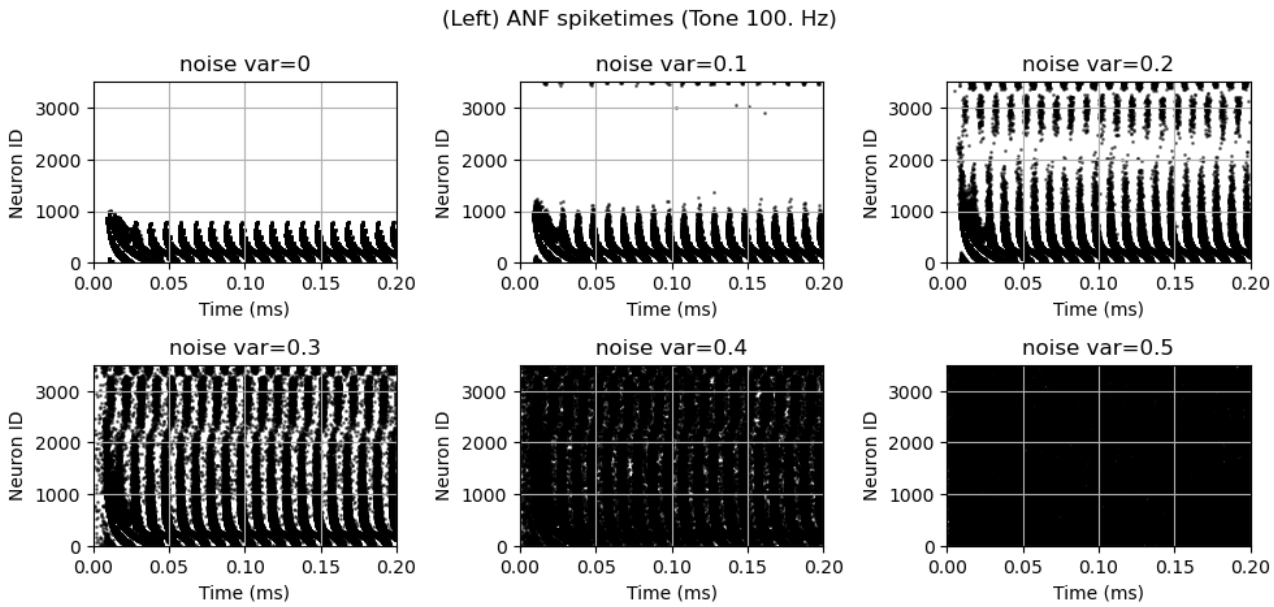


Figure 11: ANF responses with increasing noise variance.

3.1.3 Tan Carney

The processing of the Tan Carney cochlea traces the steps of the gammatone-based cochlea, with some changes. The structure follows the original Tan Carney paper [58]:

1. *HRTF*.
2. *Resampling*: Because of how the Tan Carney cochlea is implemented, it requires inputs with a sample rate of 50 kHz, but HRTFs use a more standard 44.1 kHz sample rate. So, sounds are first processed through the unchanged HRTF and then resampled to match the requirements of the following steps.
3. *CF spacing*.
4. *Middle Ear*: To model the gain introduced by the middle ear, the original paper uses a linear bandpass filter based on middle ear frequency response reported by Rosowsky [47].
5. *Cochlea*: After this, the Tan Carney cochlea substitutes the gammatone filterbank: it uses two paths, a time-varying bandpass filter for the signal and a nonlinear control path to account for compressive nonlinearity in the BM.
6. *IHC Synapse*: Finally, the IHC and the IHC-ANF synapse are modeled using the Zhang [71] model, featuring a saturating low-pass filter for the IHC and the three-store model we saw in Paragraph 1.4.1.

It is important to note that this model was not initially planned for the human cochlea: the data used to fit the parameters of the signal path mentioned in 5 used the available data of low-frequency cat ANF. Still, the low availability of high-CF cat ANFs is irrelevant for human application, as human hearing tops out around 20 kHz. As the model aims to reproduce responses from ANFs with high spontaneous rates, we do not need to add any additional noise [58].

3.1.4 Pulse Packets

The last cochlea we designed is at the opposite end of bio-plausibility and is handmade to have the cleanest possible inputs to our neural processing section. It serves as our control and shows the limit of what our neural processing section can produce with perfect inputs. It uses some basic features of NEST Simulator, the simulator for our neural processing section, which we introduce more completely in Section 3.2.1.
CFs are distributed logarithmically. We use pulse packet generators to produce a spike train with Gaussian pulse packets centered around the exact phase-locking time. A Gaussian pulsepacket is a set number of spikes (depending on the ILD), with normal distributed random displacements from the center time. The standard deviation of spike times within each pulse is a simulation parameter. We obtain two series of pulse packet generators, one series for each ear; between them, center times differ by an ITD determined from the angle with the formula of

```
delta_x = w_head * np.sin(np.deg2rad(angle))
itd = np.round(1000 * delta_x/v_sound, 2) # ms
```

To ensure proper signal propagation, we involve 10 ANFs before and 10 ANFs after the neuron with each CF, whose activity (spikes per pulse) is modulated by a Gaussian profile of amplitudes. To maintain the structure of the other peripheral models, the simulation using these pulse packet generators is run, and the resulting spike times are saved. They will then be loaded as inputs to the neural processing section.

3.2. Neural Processing

Once the peripheral model has determined the ANF spiking patterns, they are saved into a binary file to avoid re-generation for every run. Then, spike generators are created using the ANF spiking patterns, which will be the inputs for the neural processing section.

3.2.1 Simulator

For the neural processing section, we decided to use NEST Simulator [16]. Like Brian 2, NEST allows the modeling of Spiking Neural Networks, enabling experiments to be performed *in silico*. We interacted with the NEST simulator using PyNEST. This Python interface serves as an API layer translating Python requests into code made for a Simulation Language Interpreter (SLI), which finally controls the Nest Kernel, written in C++. One significant advantage of NEST was the availability of built-in thread parallelism, which enabled a substantial speedup. All simulations were run on a personal laptop, using an AMD Ryzen™ 7 7840U (up to 5.1GHz, 8-core/16-thread) with 32GB DDR5 RAM memory[17], on OpenSUSE Tumbleweed.

3.2.2 Network Structure

Because the model was inherited from a previous work [50], we will only outline the network structure and highlight the changes that were made to improve its bio-plausibility.

Cell Type	Model	Convergence	Numerosity
ANFs	parrot_neuron (devices that reproduce peripheral model output)	4 : 1 to SBCs 20 : 1 to GBCs	35000
SBCs	iaf_cond_alpha	5 : 1 to LSO PCs 5 : 1 to MSO PCs	8750
GBCs	iaf_cond_alpha	1 : 1 to LNTBC 1 : 1 to MNTBC	1750
LNTB	iaf_cond_alpha	1 : 1 to MSO PCs	1750
MNTB	iaf_cond_alpha	1 : 1 to LSO PCs 1 : 1 to MSO PCs	1750
MSO	iaf_cond_beta	1 : 1 to IC	1750
LSO	iaf_cond_alpha	1 : 1 to IC	1750
IC	iaf_cond_alpha	-	1750

Membrane time constant An important step for realism was moving to a more realistic membrane time constant (τ_m). The membrane time constant is an important property of neurons, which measures how quickly the membrane potential changes to respond to input currents and decay back to rest potential. In our application, the membrane time constant is especially relevant: as mentioned in Section 1, all cells involved in the auditory pathway have exceptionally fast time constants. In NEST, the membrane time constant is not a direct parameter but is determined from membrane capacity (C_m) and leak conductance (g_L): $\tau = C_m/g_L$. The existing model, to ensure sufficiently quick membrane time constants, used extremely low membrane capacities (1 pF), which lead to an unrealistic membrane time constant (= 0.06 ms). A realistic range [7] of C_m for bushy cells is 2 ms to 0.6 ms. This short time constant allowed for very fast response to inputs, precise temporal processing, and great ability to track input changes, but meant that (1) Bushy cell populations were able to phase lock to excessively high frequencies; (2) As soon as noisier, more realistic cochleas were used, the temporal window for coincident spikes, which triggered spiking, immediately became too small.
We moved to a higher membrane capacitance and a higher (but lower proportionally) leak conductance to increase the membrane time constant to a reasonable range. The high leak conductance simulates the fast

[17]The RAM was crucial, as simulations could take as much as 15GB of RAM

ion channels present throughout the auditory pathway. Table 1 collects the membrane capacitance and leak conductance.

Population	Membrane Capacitance (C_m, pF)	Leak Conductance (g_L, nS)
SBC	15	40
GBC	15	25
MNTBC	15	25
LNTBC	15	25
MSO	20	80
LSO	30	20
ICC	20	20

Table 1: Membrane capacitance (C_m) and leak conductance (g_L) values for the different neuron populations.

A side effect of increasing the membrane time constant was that pulse-packet generators were unable to generate spikes in the MSO.

Timing As mentioned in (2), Myoga et al. in [41] propose one plausible theory for how internal delays are generated, which does not make use of delay lines organized by frequency or cochlear delays. By delivering inhibition modeled on recorded IPSP to Mongolian gerbil brain slices, they found that inhibition dynamically shifts the timing at which excitation reaches its maximum voltage. In turn, when using bilateral inhibition and bilateral excitation, as with the bipolar MSO neuron, the coincidence detection feature of MSO neurons is modulated, allowing the researchers to shift the best ITD from zero to a range of $\pm 200\,\mu s$. To use this result, we first verified its applicability to the simple iaf_cond_beta model; in addition to the coincidence detection tuning, we also replicated its efficacy on higher frequencies. Once we confirmed its applicability, we included these findings in our model, using delays that account for the delay due to synaptic integration. To maximize contralateral bias, we used the largest delays shown in [41], $\Delta T_{inhi} = 0.2$ (difference in time of arrival between ipsilateral excitation and ipsilateral inhibition) and $\Delta T_{inhc} = -0.4$ (difference in time of arrival between ipsilateral excitation and ipsilateral inhibition):

Source pop	Dest pop	Formula	Delay (ms)
GBCs	MNTBC	–	0.45
GBCs	LNTBC	–	0.45
SBCs	MSO (ipsilateral)	–	2
SBCs	MSO (contralateral)	–	2
LNTB	MSO	$1.44 + \Delta T_{inhi}$	1.64
MNTB	MSO	$1.44 + \Delta T_{inhc}$	1.04

Table 2: Timings of connections between populations

We also verified Myoga's values in our complete network, shown in Figure 12. In the top plot, the MSO responds maximally to contralateral sounds, while in the bottom plot, the MSO responds maximally to ipsilateral sounds. All other parameters were kept constant. The middle plot shows a modest effect from a slight delay change.

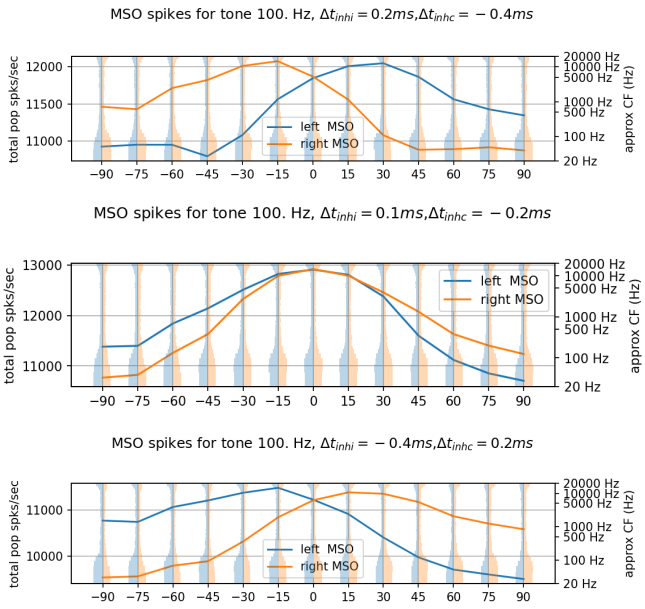


Figure 12: The effect of changing relative arrival times of inhibition from the two sides relative to the excitation arrival times.

Weights In Table 3, we show the weights used for all synaptic connections in the network.

Source population	Destination population	Synaptic Weight
ANFs	SBCs	35.0
ANFs	GBCs	7.0
GBCs	LNTBCs	20.0
GBCs	MNTBCs	30.0
SBCs	LSO	10.0
MNTBCs	LSO	-10.0
SBCs	MSO	9.0
MNTBCs	MSO	-40.0
LNTBCs	MSO	-40.0
MSO	IC	20.0
LSO	IC	20.0

Table 3: Synaptic weights between populations

3.3. Testing

Due to the multilevel aim of this work, testing was done at multiple levels. Here, we will show what our testing procedure was for each level of processing, highlighting the most relevant metrics.

3.3.1 Peripheral processing

In the peripheral section, we evaluated results at two points: first, the effect of HRTFs on sound, and then the quality and bio plausibility of ANF spikes by each of our cochleas.

HRTF To evaluate the effect of HRTFs on sounds, rather than plotting the response or the spectrum of the resulting binaural sound, we cast its effect into what our investigation was based on so far: interaural cues. This was especially relevant because of the azimuthal focus of our work. Since we preferred examining tones, as these allowed us to investigate a narrow band of neuronal fibers, ILD was measured as the maximum of the difference in the spectrum of right and left sounds. Because of noise and recording artifacts, measuring ITD needed to account for minor variations before the onset of the actual transformed sound, so we profiled the silence and

only considered the beginning of the sound from the first sampled value that exceeded a threshold proportional to the silence profile. Due to the strong impact of different HRTFs, this plot drove many of our observations.

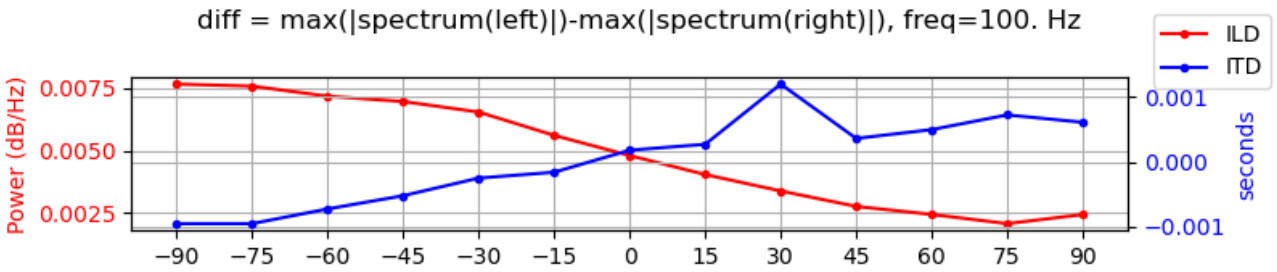


Figure 13: An example of an ITD-ILD plot from the result of an HRTF

ANF Due to the differences in our peripheral processing implementations, we decided to directly evaluate the complete effect of each on the ANF spike trains. To measure the phase-locking ability of ANFs, we used vector strength (VS), a measure of phase synchrony, or how well spike timings are synchronized to a specific phase in each period. Measuring VS across the frequency spectrum is relevant to investigate how VS changes depending on the distance from the ANF characteristic frequency. Spike rate is also an important feature of ANF: since we are not modeling different rates of ANF, we expect an average spike rate of 50 to 100 Hz at CF.

3.3.2 Neural processing

As all neural layers use the same simulation, we evaluate them together. As we needed to compare results among different trials to verify whether the network could distinguish between sounds from different angles, we needed a metric for the overall evaluation. We considered two primary metrics: average spike rate for active neurons and total population spike rate. Comparisons between different populations are not very meaningful, as due to convergence, different populations have different neuron counts. In addition, the average spike rate of the population does not introduce any additional information but only scales the total population spike rate by the population size; instead, to account for different sizes, we scaled the population spikes by the number of active neurons (i.e., neurons that spiked at least once during the observation period). Although this gave additional insight, it presented an additional issue: ANFs have a high spontaneous rate, which causes them to spike across the entire frequency range. However, the filtering capability of bushy cells caused much higher rates in bushy cell layers. So, we decided that the comparatively simpler metric of the total population spike rate was used. In addition, to evaluate how different regions of each population contributed to the overall spiking pattern, we included vertical histograms for each angle, as shown in 14.

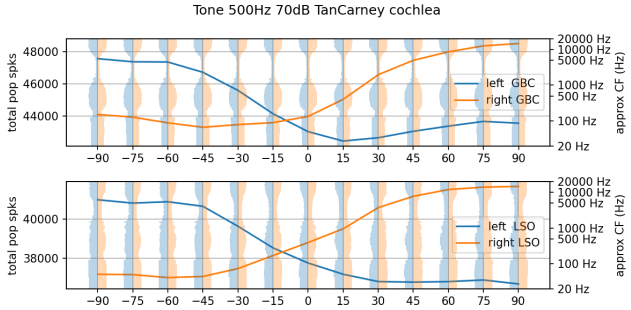


Figure 14: An example of what a result plot looks like. On the left Y-axis, the total population spikes are shown. On the right vertical axis, we show the CF corresponding to every vertical histogram bin

IC Our last objective is to evaluate a possible way to combine cue processing by the LSO and the MSO: drawing a simple excitatory connection from both of them (as seen in 1.4.3, the IC receives inputs from ipsilateral MSO and contralateral LSO), and evaluating if the result improves compared to the LSO alone. To measure the *improvement*, we fit the difference of total population spikes for each azimuthal location to a sigmoid and defined three metrics. The general-form sigmoid function we used is the following:

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

So, our metrics will be the values that minimize non-linear least squares when applying the sigmoid to our data. The correspondence is as follows:

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

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

3.4. Computational considerations

As mentioned in 3.2.1, the only hardware used for this work was a personal laptop. Because of this, computational efficiency and the ability to run multiple trials unattended were essential factors for the project's feasibility. This lead us to an important choice: NEST includes built-in thread parallelism but does not allow for process-level parallelism, as the pyNEST library converts all calls to LSI calls to the C++ kernel; inversely, brian2 does not include any support for thread parallelism, but since it is pure python it is easily process-parallelized (using multiprocessing or joblib). There is an additional, fundamental factor in our computation: the simulation requires a relatively large amount of RAM, ranging between 10 GB to 15 GB. This makes process-level parallelism infeasible, as even two processes running simultaneously would cause thrashing. At the same time, only using thread-level parallelism is not the best solution: (1) each neural simulation depends on the result of a previous, nonparallel peripheral simulation; (2) neural simulations, when run for our average time of 200 ms, only take an average of 22 s to complete; (3) peripheral simulations are recalculated on every run, even though peripheral parameters rarely change.

Because of this, we adapted to the limitations that our libraries imposed and formed the following workflow:

1. Peripheral sections are cached. Initially, we opted for a faster, hand-built caching solution based on pickle. In 5.5, we show a simplified overview of how we initially implemented it. As the code grew more complicated, more parameter types were included, and more cochleas were implemented, we decided to move to a standard implementation using joblib.Memory for disk-based caching. Although this resulted in a 10x slowdown in cache hits (from 0.0177 s to 0.190 s), it also included cache invalidation for changes to function code and more robust parameter serialization.
2. The neural section uses thread-level parallelism. NEST handles parallelism quite well, and by tuning the thread count to our processors' (16), we obtained a $\tilde{8}$X speedup.
3. When making significant changes to a peripheral section, or testing something that requires new generation of an ANFResponse, we use a simple multiprocessing based script to generate and cache all ANF responses so that the following simulation runs would only load them and not generate them. This became progressively more relevant as the complexity of our peripheral sections grew[18].

This workflow allowed us to maximize efficiency, by parallelizing both sections of our codebase, using the most efficient version for each.

[18] Generating ANF spiketrains for our most bioplausible took over a minute for each side, per trial.

4. Results

In this section, we report our findings. As for all previous sections, we maintain the sequential structure, starting with peripheral processing and progressively approaching higher-level nuclei.

4.1. Peripheral processing

The first approach used for objective **??** was modeling the processing happening outside of the ear, by using human HRTFs for realistic sound transformations. Then, we implemented one synthetic cochlea and two realistic peripheral processing pathways. The synthetic cochlea (1) was based on Pulse Packet generators (a generator type in NEST Simulator which allows to set spike times manually); the realistic peripheral processing pathways used (2) a gammatone-based cochlea and (3) the Tan Carney cochlea paired with earlier models of middle ear amplifier and IHC-ANF synapses.

4.1.1 HRTF

The effect of the head-related transfer functions was larger than we expected. Figure 16 shows the effect of HRTF on ILD and ITD as we change the subject on which the HRTF was based (left) and the frequency of the tone (right). For what regards changing the subject, the difference was felt much more strongly in ILD curves: while some (top-left, both images bottom row) show ILD growth at -75°,-60°compared to -90°before decreasing, others (top-right, center left) immediately decrease; at the other end, most responses show a non-monotonic decrease, with the final +90°varying in magnitude; there is a strong asymmetry shown in most responses, with the response never crossing the x-axis in correspondence with zero difference, as in Figure 15.

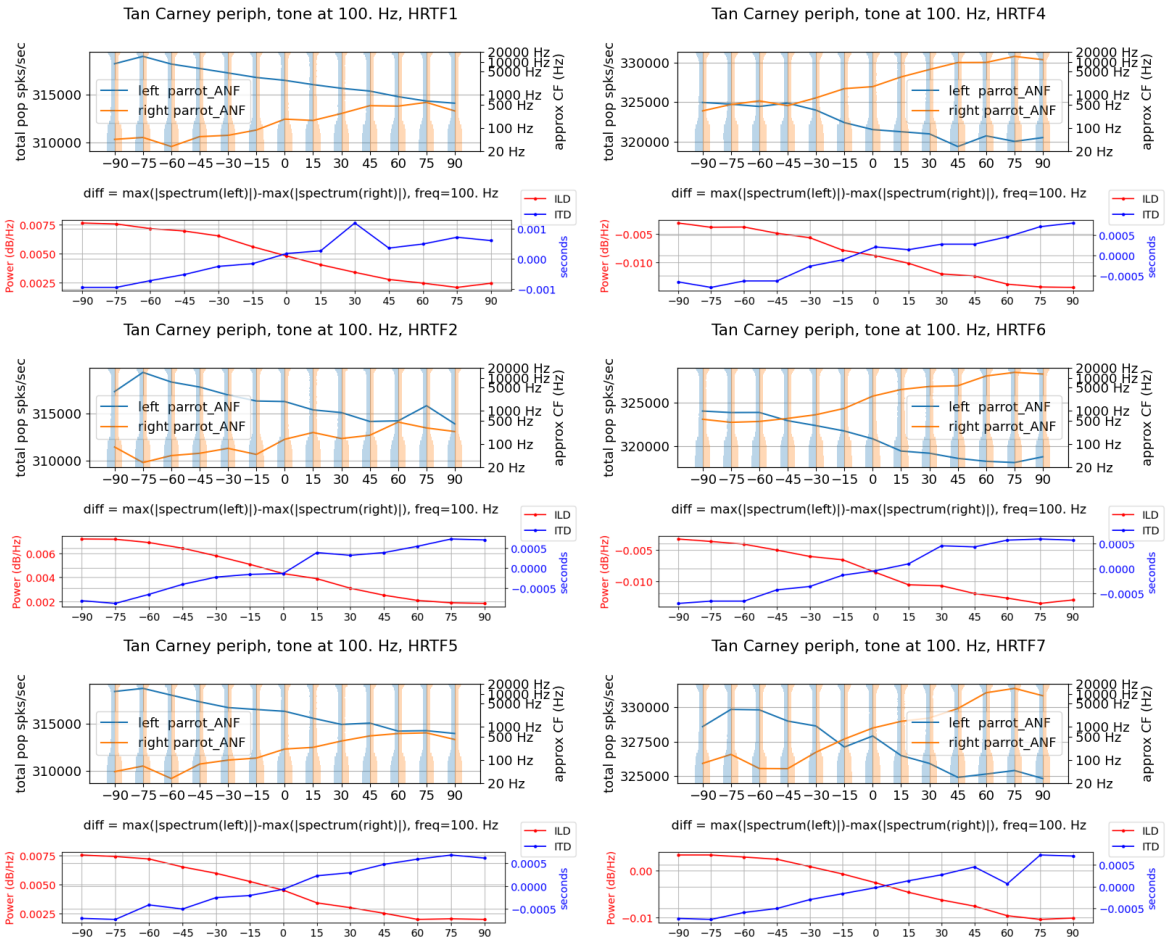


Figure 15: ANF responses for different HRTFs. In the left column, HRTFs where the left ear always receives a higher level than the right ear, and in the right column, HRTFs where the right ear mostly receives a higher level than the left ear.

The results in responses to different frequencies are as expected, as ILD size increases as frequency increases,

while ITD remains constant. ITDs show some noise in their responses.

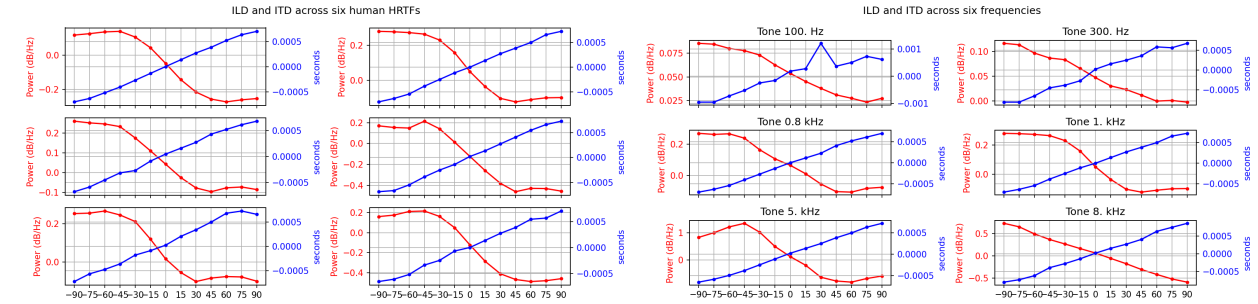


Figure 16: The effects on measured ITD and ILD for varying human HRTF (left) and varying frequencies (right). The six human HRTFs were tested using a 1 kHz tone.

4.1.2 Non neural processing

As mentioned in 3.3.1, we evaluate peripheral processing pathways based on the ANF response, focusing on vector strength (VS). Our results are collected in 17.

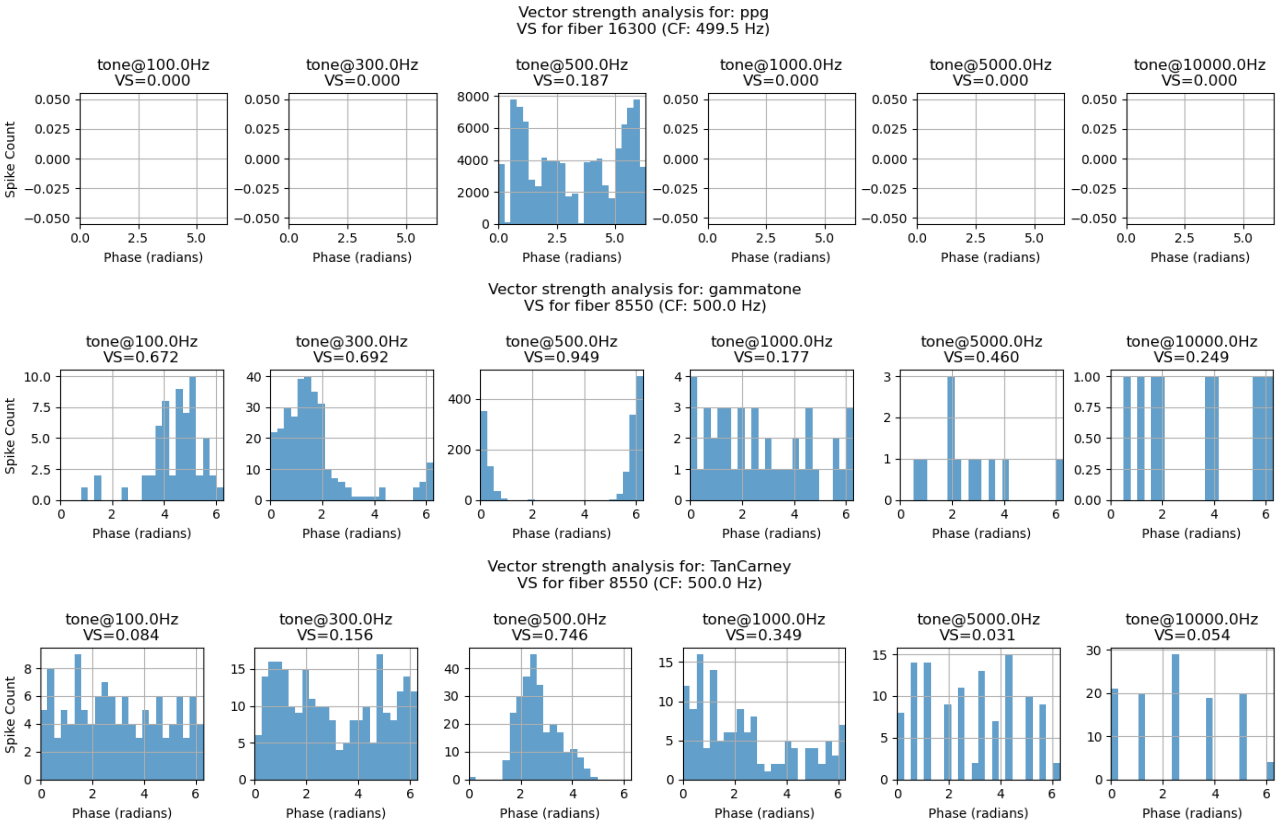


Figure 17: Vector strength plots for the three peripheral sections

In the figure we plot an analysis of the spiking behavior of a single auditory nerve fiber, one of the ten connected to the IHC, with the characteristic frequency listed above. The inputs were three-second tones, with the frequencies listed above each column of plots. The results show that using more biologically plausible peripheral processing results in higher-quality ANF spike trains. We highlight the main findings here: (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), which is in disagreement with the more bioplausible options (2) the phase locking exhibited by the synthetic processing was very limited (3) the gammatone-based processing shows very accurate phase locking for ANFs whose CFs were close to the input frequency, with a slow falloff in ANFs further away from CF (4) 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 shown (5) the most bioplausible processing shows comparatively low spiking rates, which fall off slowly in ANFs with

different CFs, and (6) lower CF ANFs are still active even if the CF is further away. These findings show the superiority of a realistic, bioplausible processing: 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 CF changes.

4.2. Neural processing

Once ANF spiketrains were generated (and cached), they became the input to our neural simulation. As mentioned in 3.3.2, both total population spike rate and active neuron spike rate are shown as population spiking metrics, depending on what the focus is. In this section, we will analyze how information flowed to each population.

4.2.1 Intermediate populations

The role of bushy cells is to filter the ANF input. In 18, we show the raster plot of spiking behavior from a range of neurons surrounding the cells with CF closest to the input tone (here, 100 Hz). The corresponding index to CF in ANFs was determined from the rule that every IHC has 10 ANFs attached to it, while the other was calculated from the convergence rules. The striped patterns shown are a feature of the tonotopic organization of the cochlea as CFs grow exponentially.

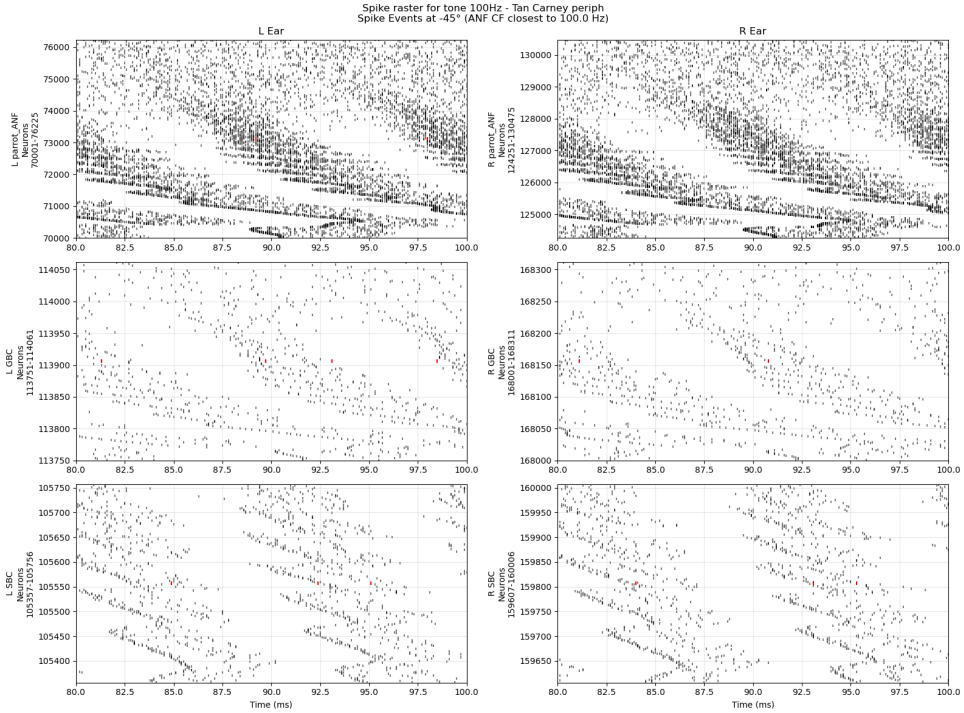


Figure 18: Raster plots of neurons surrounding CF in ANFs and bushy cell populations. Due to population size differences, ANF plots show spikes from a range much larger than bushy layers.

Following populations, cells in the LNTB and MNTB do not show any change in spiking rates from their input, the GBCs. This is consistent with the literature, as GBCs form strong and stable synapses, and the role of LNTB and MNTB is to relay their signals and faithfully transform them into inhibitory contributions.

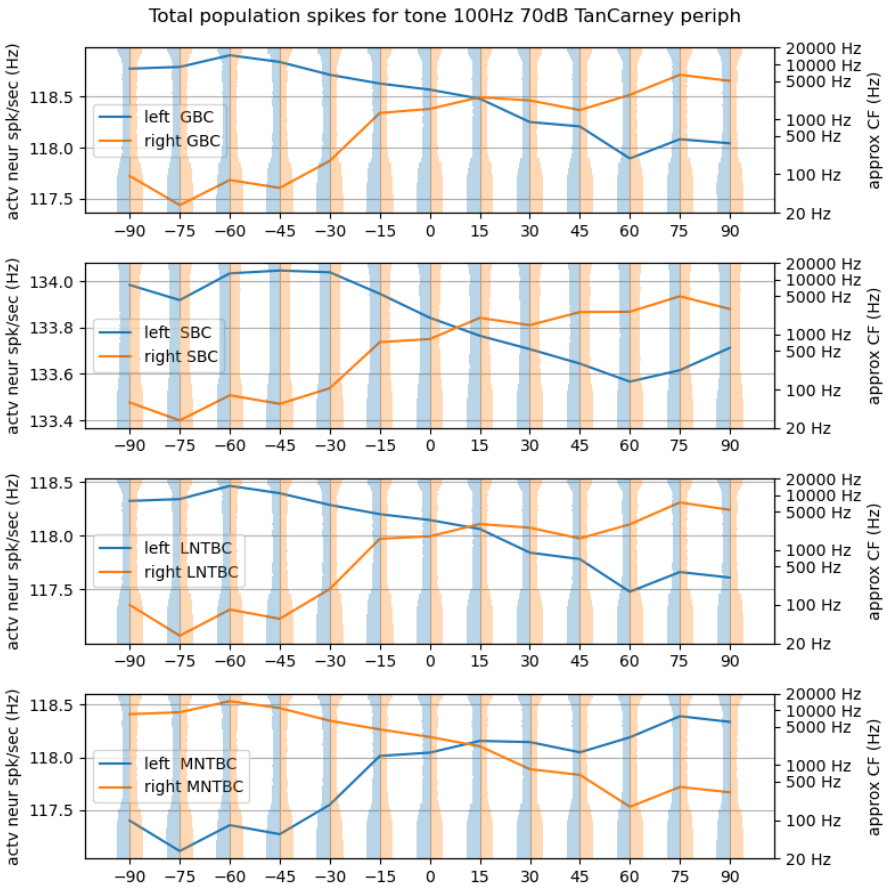


Figure 19: Population spiking rates for 100Hz tone

4.2.2 Higher centers

We now move to analyzing the spiking behaviors of nuclei in the SOC, the LSO, and the MSO.

LSO In our testing, the LSO was remarkably stable. Even at low frequencies, where the sizes of ILD are considered minute, the shape of the LSO was unmistakable and very resistant to perturbations. The LSO remained consistent when membrane time constants and weights were changed. It was also generated equally well by both realistic cochleas, as shown in 20. The difference between spike volumes of the two sides increased, as is to be expected with the rising ILD, as frequency increased. When using the gammatone-based peripheral section, the LSO failed above 10 kHz.

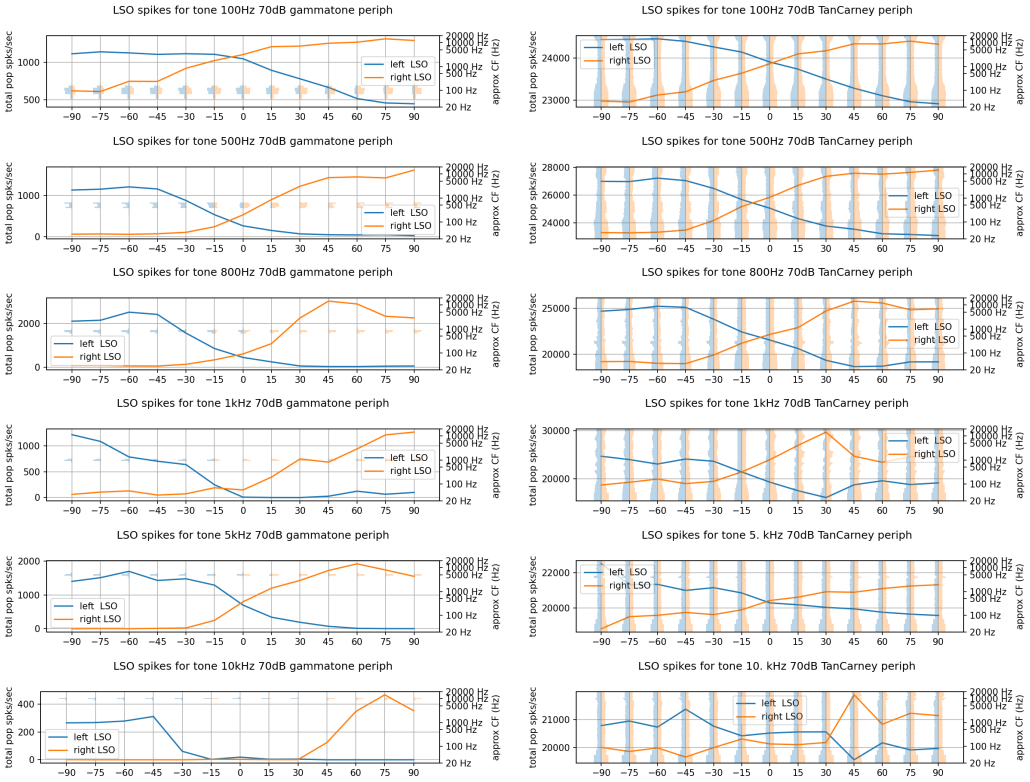


Figure 20: Population spike rates for the LSO with growing frequencies, for gammatone-based (left) and Tan Carney (right) processing.

MSO As with the LSO, the largest difference between the two peripheral processing is the size of involved frequency ranges: while the gammatone-based processing only activates a narrow portion of ANFs (and hence the rest of the network), the more bioplausible inputs generate a broader activation. At the same time, MSO cells with high CF activate much less in proportion, 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 continues to maintain efficacy (except for at 1 kHz) to 10 kHz.

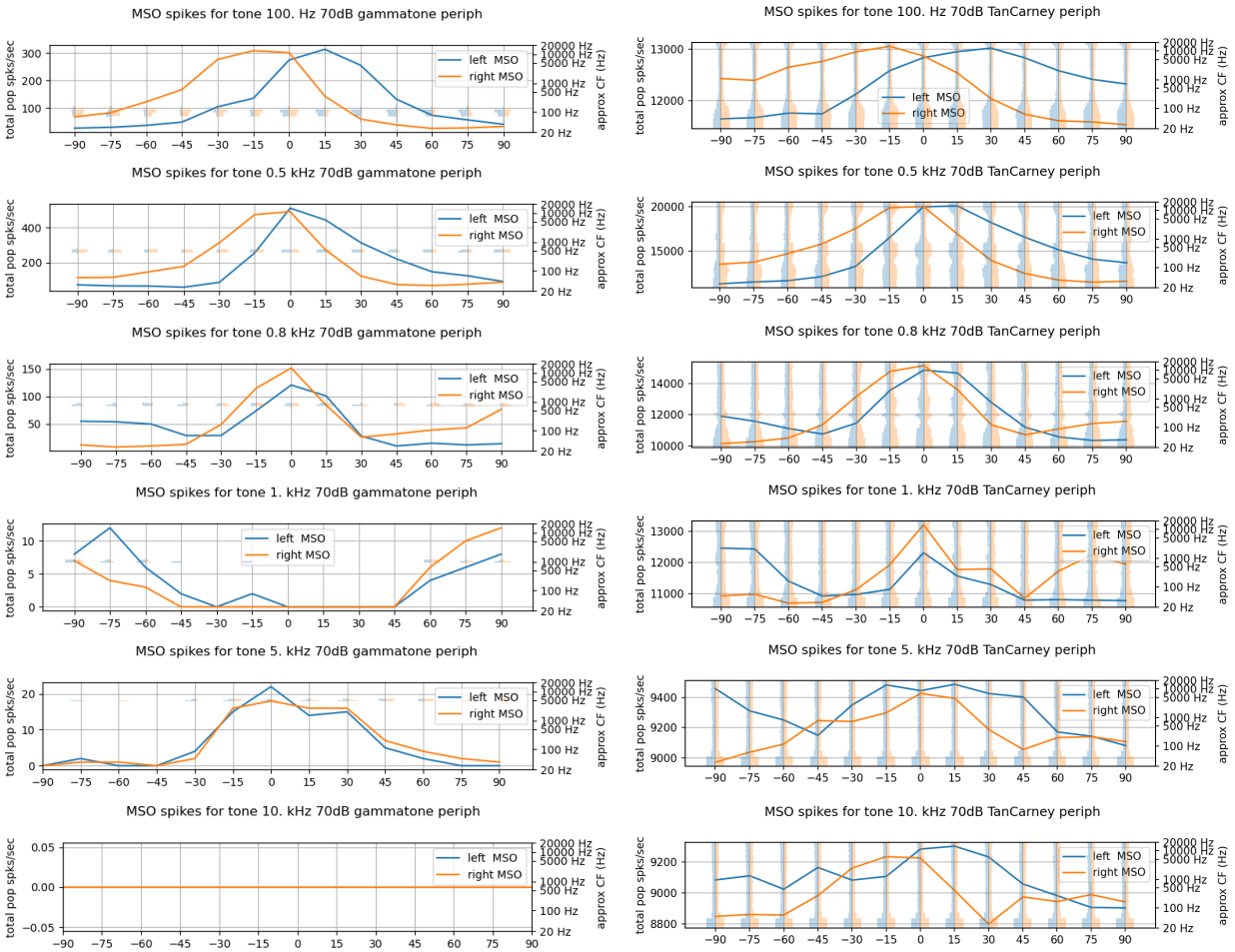


Figure 21: Population spike rates for the MSO with growing frequencies, for gammatone-based (left) and Tan Carney (right) processing.

4.3. IC and integrating cues

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. In 22, we plot the difference of total population spikes between the right and left LSO (top) and IC (bottom). All differences are collected in table 4. 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 increased 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 only by around 3 degrees, which is not a significant improvement.

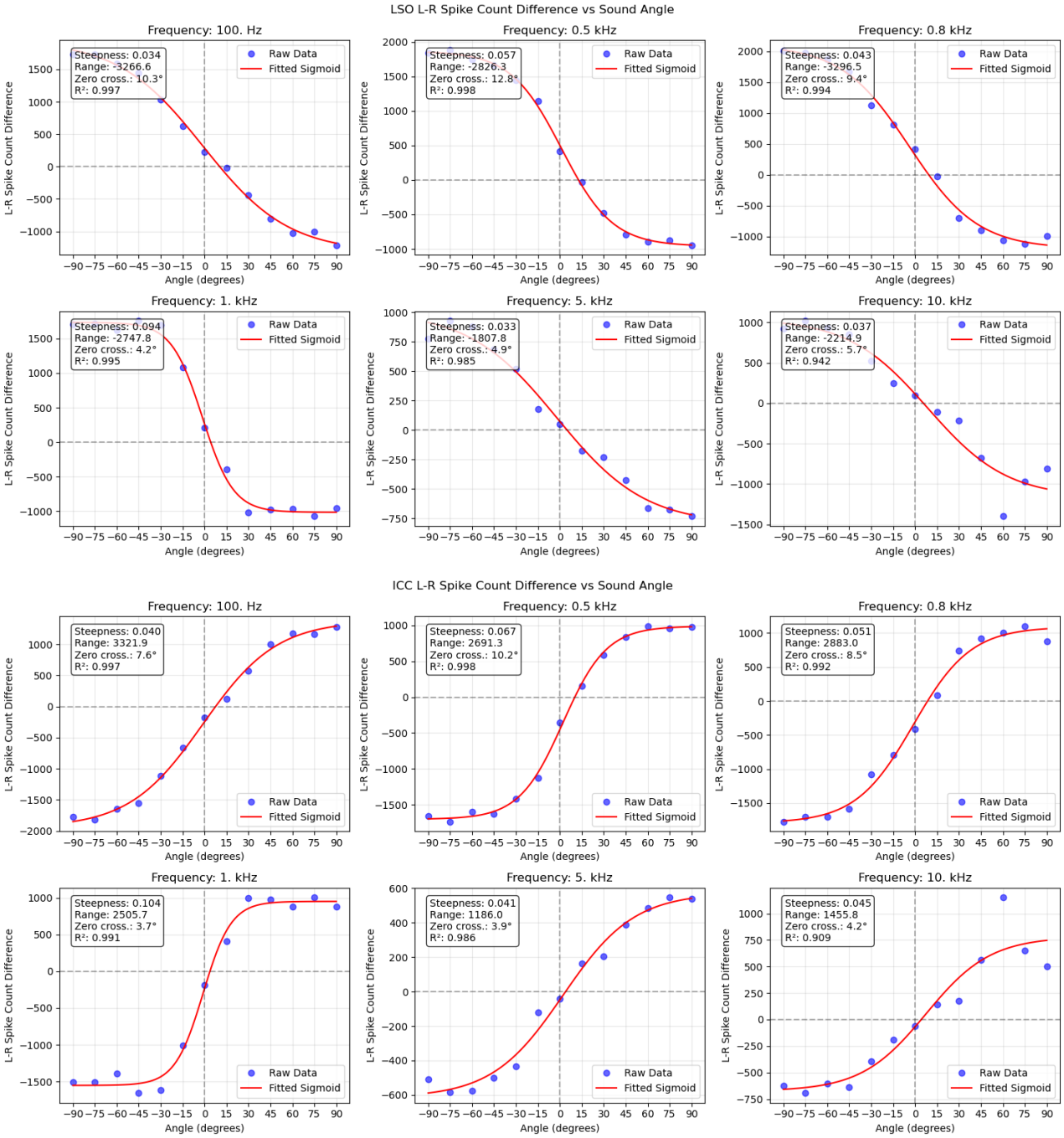


Figure 22: Comparison of sigmoid fit for all frequencies. Top: LSO. Bottom: IC. Because the contralateral LSO connects to IC, its slope is in the opposite direction.

Table 4: Analysis of improvement of ICC over LSO

STEEPNESS				
Frequency	LSO	ICC	Diff	Improvement
100	0.034	0.040	0.005	15.5%
500	0.057	0.067	0.010	16.8%
800	0.043	0.051	0.008	18.7%
1000	0.094	0.104	0.010	10.2%
5000	0.033	0.041	0.008	24.4%
10000	0.037	0.045	0.009	23.2%
Average				18.1%
R_SQUARED				
Frequency	LSO	ICC	Diff	Improvement
100	0.997	0.997	0.000	0.0%
500	0.998	0.998	-0.000	-0.0%
800	0.994	0.992	-0.002	-0.2%
1000	0.995	0.991	-0.004	-0.4%
5000	0.985	0.986	0.001	0.1%
10000	0.942	0.909	-0.033	-3.5%
Average				-0.7%
RANGE				
Frequency	LSO	ICC	Diff	Improvement
100	3266.574	3321.874	55.301	1.7%
500	2826.344	2691.311	-135.032	-4.8%
800	3296.494	2882.988	-413.506	-12.5%
1000	2747.825	2505.705	-242.119	-8.8%
5000	1807.770	1185.965	-621.805	-34.4%
10000	2214.870	1455.782	-759.088	-34.3%
Average				-15.5%
ZERO-CROSSING ACCURACY				
Frequency	LSO	ICC	Diff	Improvement
100	10.316	7.556	2.759	26.7%
500	12.788	10.226	2.562	20.0%
800	9.356	8.534	0.822	8.8%
1000	4.197	3.662	0.535	12.7%
5000	4.942	3.942	1.001	20.2%
10000	5.660	4.202	1.458	25.8%
Average				19.1%

5. Discussion

Our results show significant improvements on multiple fronts but also limits to our model. We consider some of them here.

5.1. Peripheral processing

Overall, we found significant advantages in implementing and testing bioplausible peripheral processing sections. The most impactful feature covered the unavailability of data: most bioplausible models are tested against existing recordings and features, but the original data is limited and difficult to obtain. Because of this, our more advanced processing pipeline was able to inform things like the multiplicative scalar for the gammatone-based peripheral processing. On the other hand, the Tan-Carney-based model for ANFs, because of its bioplausibility, did not produce ANFs strictly sectioned to a frequency range, which affected our parameter search: for example, when using high enough weights for inhibition, the MSO stops considering high-frequency data, and the only fibers that remain active are the ones with low CF. On the other hand, our least bioplausible ANF model did not produce useful results, as its spikes were so closely aligned that the parameter values needed to allow their processing were entirely different from both other models.
Our bioplausible models imposed two limitations on us: the 15°increments in possible azimuthal positions and the adaptation shown by the bioplausible IHC-ANF synapse due to vesicle availability. The first could be tackled by including a more modern approach to HRTF filtering, such as the python library sofar. In contrast, the second one, also true in biological experiments, can be solved by repeating the stimulus multiple times in a brief window.

5.2. Neural pathway

The intermediate populations of our neural processing resulted in modestly bioplausible results: while bushy cells were able to phase lock to low-frequency stimuli, our modeling did not include significant differences between spherical and globular bushy cells beyond their convergence values, and no individual differences among the population. Still, they accomplished their filtering task and were able to bring reliable, phase-locked PSPs to the upper nuclei.

5.3. MSO

In our modeling of the MSO, we confirm the findings of Myoga and the limitation of this approach: the maximal shift we could confirm was $\pm 200ms$, roughly equivalent to $\pm 30°$ with human head size. This would not be very compatible with a two-channel approach, as it would only be able to determine sound location in a very narrow band of locations in front of the subject; this is not compatible with behavioral evidence. It also does not offer sufficient range to provide a map of ITDs. At the same time, while experimenting with this approach, we learned that it may be possible to use the relative strength of the two inhibition sources to drive the angle of maximal response. A preliminary finding is shown in 23.

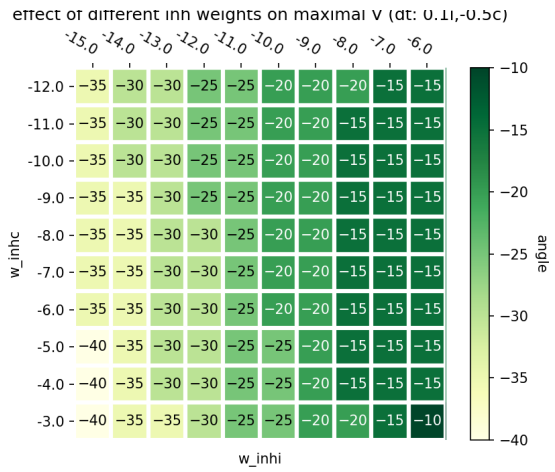


Figure 23: Colormap showing which angle (hence ITD) makes MSO reach maximum voltage when varying strength of inhibition sources. Both excitation sources are kept at weight 6

5.4. LSO

As mentioned, the LSO had remarkably stable outputs, which reliably showed differences (modest at low frequencies, larger at high frequencies) between sides. The modest difference at low frequencies is also due to a limited sensibility to ITD from the LSO: in Figure 24 we show the spike behavior of the LSO with an input tone filtered through an artificial HRTF which only includes ITD information and no spectral (hence no ILD) differences. A slight sensitivity to ITD by LSO is also documented in literature [68].

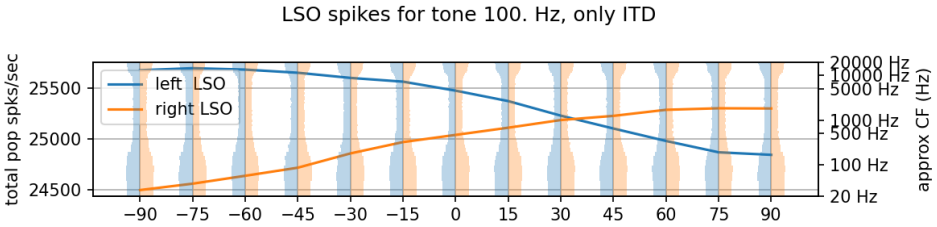


Figure 24: LSO response to a 100Hz tone filtered through an artificial, ITD-only HRTF.

5.5. IC

Our simplified, EE-designed IC showed fairly accurate results, but these were mostly based on the LSO results, as was shown by our analysis of IC vs LSO spiking behaviors. The loss in range was expected, as the MSO spiking curve shows peaks around zero degrees, while the LSO curve is highest at the tails. In our experiments, we also attempted to adjust network features to decrease LSO spiking rates and increase MSO rates artificially. Still, due to the diversity of involved populations (high frequency for LSO and low frequency for MSO), the IC remained mainly in agreement with the LSO.

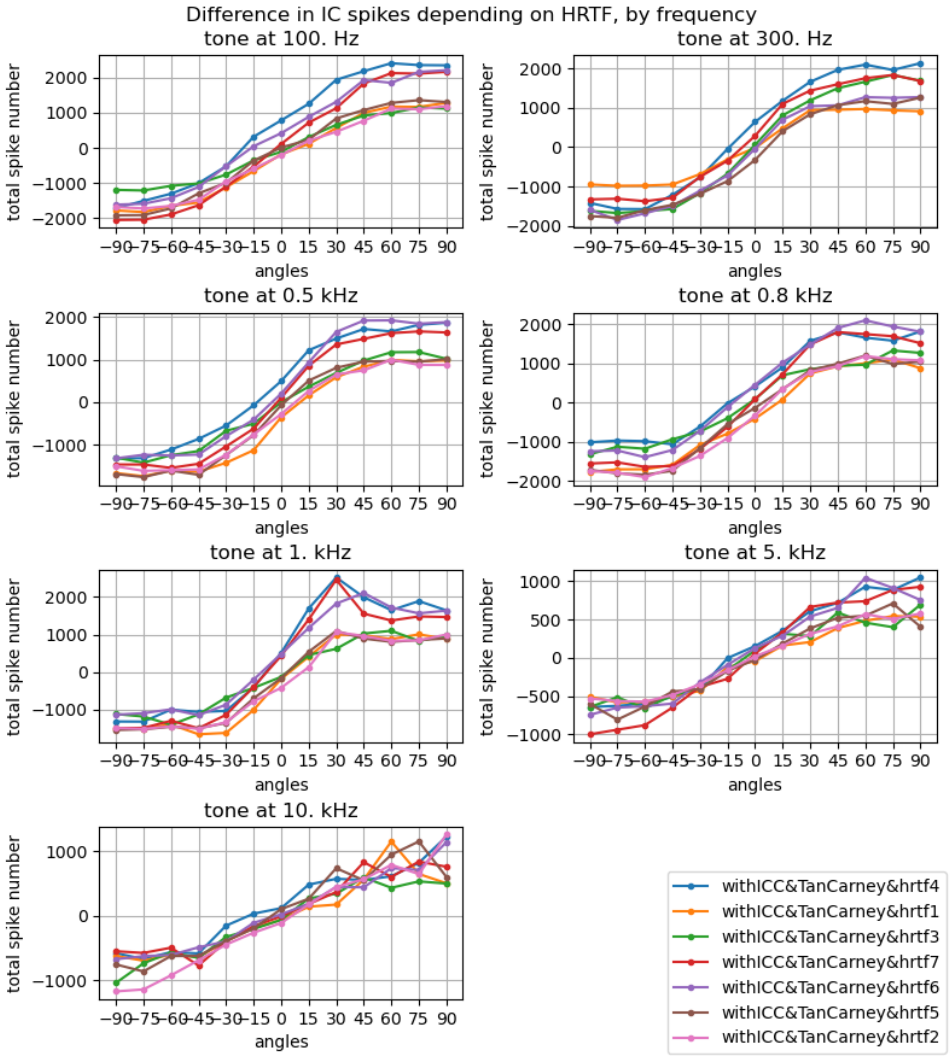


Figure 25: Variation in side-to-side difference of IC responses depending on HRTF, by frequency

The IC response curves were not excessively affected by using different seeds for the random realizations of the peripheral processing (Figure 26). HRTF had a larger impact (Figure 25), and in limited tests, the results were positive for various omnidirectional white noise levels (Figure 27).
The limited impact of MSO on IC curves is in agreement with literature: the majority of IC neurons show ILD sensitivity, which would suggest an LSO role in determining IC response [18], and ILDs are the strongest cue for spatial tuning in the BIC [54]. It is possible, then, that the MSO is not used for absolute localization, but instead serves as a relative, sound segregation system, as proposed in [17]. This would suggest a paradigm shift: from a simple transformation of inputs to a flexible representation of soundscapes and environments in time.

References

[1] gammatoneFilterBank - MATLAB documentation. Source: https://www.mathworks.com/help/audio/ref/gammatonefilterbank-system-object.html.

[2] Listen HRTF database. Source: http://recherche.ircam.fr/equipes/salles/listen/.

[3] Edgar F. Allin. Evolution of the mammalian middle ear. 147(4):403–437. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmor.1051470404.

[4] Jonathan Ashmore. Cochlear outer hair cell motility. 88(1):173–210.

[5] Antje Brand, Oliver Behrend, Torsten Marquardt, David McAlpine, and Benedikt Grothe. Precise inhibition is essential for microsecond interaural time difference coding. 417(6888):543–547. Publisher: Nature Publishing Group.

[6] Douglas S. Brungart and William M. Rabinowitz. Auditory localization of nearby sources. head-related transfer functions. 106(3):1465–1479.

[7] Xiao-Jie Cao, Shalini Shatadal, and Donata Oertel. Voltage-sensitive conductances of bushy cells of the mammalian ventral cochlear nucleus. 97(6):3961–3975.

[8] L. H. Carney, M. J. McDuffy, and I. Shekhter. Frequency glides in the impulse responses of auditory-nerve fibers. 105(4):2384–2391.

[9] Laurel H. Carney. A model for the responses of low-frequency auditory-nerve fibers in cat. 93(1):401–417.

[10] C. E. Carr and M. Konishi. A circuit for detection of interaural time differences in the brain stem of the barn owl. 10(10):3227–3246. Publisher: Society for Neuroscience Section: Articles.

[11] M. A. Cheatham and P. Dallos. Inner hair cell response patterns: implications for low-frequency hearing. 110(4):2034–2044.

[12] J. A. Clack. The evolution of tetrapod ears and the fossil record. 50(4):198–212.

[13] 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/.

[14] Tom P. Franken, Philip H. Smith, and Philip X. Joris. In vivo whole-cell recordings combined with electron microscopy reveal unexpected morphological and physiological properties in the lateral nucleus of the trapezoid body in the auditory brainstem. 10:69.

[15] Rong Z. Gan, Qunli Sun, Robert K. Dyer, Kuang-Hua Chang, and Kenneth J. Dormer. Three-dimensional modeling of middle ear biomechanics and its applications. 23(3):271–280.

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

[17] Benedikt Grothe and Michael Pecka. The natural history of sound localization in mammals - a story of neuronal inhibition. 8.

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

[19] John J. Guinan, Shelley S. Guinan, and Barbara E. Norris. Single auditory units in the superior olivary complex: I: Responses to sounds and classifications based on physiological properties. 4(3):101–120. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/00207457209147165.

[20] Ruxue Guo, Ruiyu Liang, Qingyun Wang, and Cairong Zou. A design method for gammachirp filterbank for loudness compensation in hearing aids. 12(4):1793. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[21] Nicol S. Harper and David McAlpine. Optimal neural population coding of an auditory spatial cue. 430(7000):682–686. Publisher: Nature Publishing Group.

[22] P. A. Hill, P. A. Nelson, O. Kirkeby, and H. Hamada. Resolution of front-back confusion in virtual acoustic imaging systems. 108(6):2901–2910.

[23] Stephen D. Holmes, Christian J. Sumner, Lowel P. O'Mard, and Ray Meddis. The temporal representation of speech in a nonlinear model of the guinea pig cochlea. 116(6):3534–3545.

[24] Kathryn Hopkins. Chapter 27 - deafness in cochlear and auditory nerve disorders. In Michael J. Aminoff, François Boller, and Dick F. Swaab, editors, *Handbook of Clinical Neurology*, volume 129 of *The Human Auditory System*, pages 479–494. Elsevier.

[25] T. Irino and R. D. Patterson. A compressive gammachirp auditory filter for both physiological and psychophysical data. 109(5):2008–2022.

[26] Eric Javel. Auditory system, peripheral. In Michael J. Aminoff and Robert B. Daroff, editors, *Encyclopedia of the Neurological Sciences*, pages 305–311. Academic Press.

[27] P. X. Joris, L. H. Carney, P. H. Smith, and T. C. Yin. Enhancement of neural synchronization in the anteroventral cochlear nucleus. i. responses to tones at the characteristic frequency. 71(3):1022–1036.

[28] P. X. Joris, L. H. Carney, P. H. Smith, and T. C. Yin. Enhancement of neural synchronization in the anteroventral cochlear nucleus. i. responses to tones at the characteristic frequency. 71(3):1022–1036.

[29] Shotaro Karino, Philip H. Smith, Tom C. T. Yin, and Philip X. Joris. Axonal branching patterns as sources of delay in the mammalian auditory brainstem: a re-examination. 31(8):3016–3031.

[30] Sukant Khurana, Michiel W. H. Remme, John Rinzel, and Nace L. Golding. Dynamic interaction of ih and IK-LVA during trains of synaptic potentials in principal neurons of the medial superior olive. 31(24):8936–8947.

[31] Abhijit Kulkarni and H. Steven Colburn. Infinite-impulse-response models of the head-related transfer function. 115:1714–28.

[32] Enrique A. Lopez-Poveda, Luis F. Barrios, and Ana Alves-Pinto. Psychophysical estimates of level-dependent best-frequency shifts in the apical region of the human basilar membranea). 121(6):3646–3654.

[33] Enrique A. Lopez-Poveda and Almudena Eustaquio-Martín. A biophysical model of the inner hair cell: the contribution of potassium currents to peripheral auditory compression. 7(3):218–235.

[34] Enrique A. Lopez-Poveda and Ray Meddis. A human nonlinear cochlear filterbank. 110(6):3107–3118.

[35] Enrique A. Lopez-Poveda. Spectral processing by the peripheral auditory system: Facts and models. In *International Review of Neurobiology*, volume 70 of *Auditory Spectral Processing*, pages 7–48. Academic Press.

[36] Geoffrey A. Manley. An evolutionary perspective on middle ears. 263(1):3–8.

[37] D. McAlpine, D. Jiang, and A. R. Palmer. A neural code for low-frequency sound localization in mammals. 4(4):396–401.

[38] David McAlpine, Dan Jiang, and Alan R. Palmer. A neural code for low-frequency sound localization in mammals. 4(4):396–401. Publisher: Nature Publishing Group.

[39] Ray Meddis. Auditory-nerve first-spike latency and auditory absolute threshold: a computer model. 119(1):406–417.

[40] Ray Meddis and Enrique Lopez-Poveda. Auditory periphery: From pinna to auditory nerve. volume 35, pages 7–38.

[41] 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.

[42] Brad Norton. What is valorant's HRTF setting? audio enhancements explained.

[43] Russell R. Pfeiffer. Classification of response patterns of spike discharges for units in the cochlear nucleus: Tone-burst stimulation. 1(3):220–235.

[44] Christopher Plack, Andrew Oxenham, and Vit Drga. Linear and nonlinear processes in temporal masking. 88:348–358.

[45] Lord Rayleigh. XII. on our perception of sound direction. 13(74):214–232. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14786440709463595.

[46] Michael T. Roberts, Stephanie C. Seeman, and Nace L. Golding. A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry. 78(5):923–935.

[47] John J. Rosowski. Models of external- and middle-ear function. In Harold L. Hawkins, Teresa A. McMullen, Arthur N. Popper, and Richard R. Fay, editors, *Auditory Computation*, pages 15–61. Springer.

[48] M. A. Ruggero and N. C. Rich. Furosemide alters organ of corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. 11(4):1057–1067. Publisher: Society for Neuroscience Section: Articles.

[49] Mario A. Ruggero and Andrei N. Temchin. Middle-ear transmission in humans: wide-band, not frequency-tuned? 4:53–58.

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

[51] Luisa L. Scott, Paul J. Mathews, and Nace L. Golding. Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive. 25(35):7887.

[52] Chandran V. Seshagiri and Bertrand Delgutte. Response properties of neighboring neurons in the auditory midbrain for pure-tone stimulation: a tetrode study. 98(4):2058–2073.

[53] 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.

[54] 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.

[55] H. Spoendlin and A. Schrott. Analysis of the human auditory nerve. 43(1):25–38.

[56] S. S. Stevens. On the psychophysical law. 64(3):153–181. Place: US Publisher: American Psychological Association.

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

[58] 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.

[59] Daniel J. Tollin. The lateral superior olive: A functional role in sound source localization. 9(2):127–143. Publisher: SAGE Publications Inc STM.

[60] Daniel J. Tollin and Tom C. T. Yin. The coding of spatial location by single units in the lateral superior olive of the cat. II. the determinants of spatial receptive fields in azimuth. 22(4):1468–1479.

[61] C. Tsuchitani. Functional organization of lateral cell groups of cat superior olivary complex. 40(2):296–318.

[62] C. Tsuchitani. Functional organization of lateral cell groups of cat superior olivary complex. 40(2):296–318.

[63] Marcel van der Heijden, Jeannette A.M. Lorteije, Andrius Plauška, Michael T. Roberts, Nace L. Golding, and J. Gerard G. Borst. Directional hearing by linear summation of binaural inputs at the medial superior olive. 78(5):936–948.

[64] Timothy Walsh, Leszek Demkowicz, and Richard Charles. Boundary element modeling of the external human auditory system. 115(3):1033–1043.

[65] T. F. Weiss and C. Rose. A comparison of synchronization filters in different auditory receptor organs. 33(2):175–179.

[66] Mario Wolf, Pascalis Trentsios, Niklas Kubatzki, Christoph Urbanietz, and Gerald Enzner. *Implementing Continuous-Azimuth Binaural Sound in Unity 3D*. Pages: 389.

[67] T. C. Yin and J. C. Chan. Interaural time sensitivity in medial superior olive of cat. 64(2):465–488.

[68] 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.

[69] William A. Yost. Sound source localization identification accuracy: Level and duration dependencies. 140(1):EL14.

[70] Xuedong Zhang and Laurel H. Carney. Analysis of models for the synapse between the inner hair cell and the auditory nerve. 118(3):1540–1553.

[71] Xuedong Zhang, Michael G. Heinz, Ian C. Bruce, and Laurel H. Carney. A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear tuning with compression and suppression. 109(2):648–670.

[72] E. Zwicker and E. Terhardt. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. 68(5):1523–1525.

Appendix

```
def load_anf_response(tone, angle, cochlea_key, params):
        filepath = (
        Paths.ANF_SPIKES_DIR
        / cochlea_key
        / create_sound_key(tone)
        / f"{create_ex_key(tone,deg,params)}.pic"
    )
    if os.path.isfile(filepath):
        with open(filepath, "rb") as f:
            anf: AnfResponse = pickle.load(f)
    else:
        anf = cochlea_func(tone, angle, params)
        filepath.parent.mkdir(exist_ok=True, parents=True)
        with open(filepath, "wb") as f:
            pickle.dump(anf, f)
    return anf
```

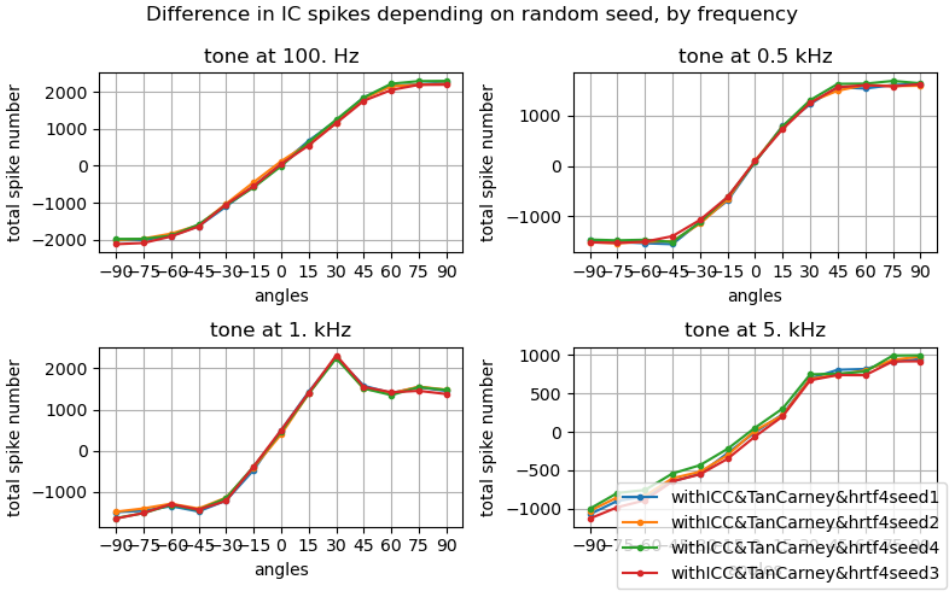


Figure 26: Variation in the side-to-side difference of IC responses depending on seed

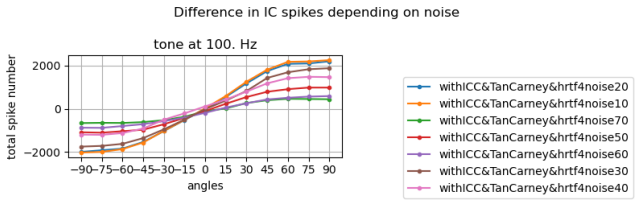


Figure 27: Variation in the side-to-side difference of IC responses depending on background, omnidirectional noise at various levels, for a 100Hz tone.

Abstract in lingua italiana

La capacità di localizzare i suoni nello spazio è uno delle funzioni più studiate dell'udito. Tuttavia, nonostante l'apparente semplicità, i meccanismi che stanno alla base di questa capacità nei mammiferi sono ancora ignoti. In questa tesi siamo partiti da una rete neuronale spiking esistente che imita il circuito del tronco encefalico uditivo e la sua organizzazione tonotopica e ne abbiamo aumentato la bioplausibilità. Abbiamo applicato un duplice focus su input e output della rete: utilizzando modelli avanzati di elaborazione periferica, abbiamo simulato come il suono viene ricevuto e tradotto in segnali neurali dalla coclea; questi segnali sono in seguito elaborati da una simulazione neurale e integrati dai nuclei superiori. Abbiamo poi analizzato le più recenti teorie in merito a questa integrazione e esplorato una possibile strategia per ottenere un'ulteriore elaborazione nel mesencefalo.

Parole chiave: neuroscienze computazionali, localizzazione uditiva, simulazione neurale, rete neuronale

Acknowledgements

This work is the result of months of effort, misteps, moments of satisfaction, bugs, and plenty of doubts. I am not afraid to say I would not have made it through all of them if it wasn't for the ever-lasting support of many people, inside and outside the lab. I must thank my co-supervisor, Francesco De Santis, for patiently guiding me through the forest of an unknown discipline; my supervisor, Prof. Alberto Antonietti, for showing me just how much there is to know in computational neuroscience (and how fun it can be to study it!); and my long list of dedicated teachers. I will forever be in debt to my girlfriend, my family and my friends for sticking with me at my worst and encouraging me to my best.