

## RESEARCH ARTICLE

## AIM: A network model of attention in auditory cortex

Kenny F. Chou<sup>1,2\*</sup>, Kamal Sen<sup>1,2,3,4\*</sup>

**1** Department of Biomedical Engineering, Boston University, Boston, Massachusetts, United States of America, **2** Hearing Research Center, Boston University, Boston, Massachusetts, United States of America, **3** Neurophotonics Center, Boston University, Boston, Massachusetts, United States of America, **4** Center for Systems Neuroscience, Boston University, Boston University, Boston, Massachusetts, United States of America

\* [kfchou@bu.edu](mailto:kfchou@bu.edu) (KC); [kamalsen@bu.edu](mailto:kamalsen@bu.edu) (KS)

## OPEN ACCESS

**Citation:** Chou KF, Sen K (2021) AIM: A network model of attention in auditory cortex. *PLoS Comput Biol* 17(8): e1009356. <https://doi.org/10.1371/journal.pcbi.1009356>

**Editor:** Boris S. Gutkin, École Normale Supérieure, Collège de France, CNRS, FRANCE

**Received:** December 22, 2020

**Accepted:** August 18, 2021

**Published:** August 27, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1009356>

**Copyright:** © 2021 Chou, Sen. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All code used for this work are available online at [www.github.com/kfchou/AIM\\_network](https://www.github.com/kfchou/AIM_network).

**Funding:** This work was supported by an NSF Award 1835270 (to KS). The funders had no role in

## Abstract

Attentional modulation of cortical networks is critical for the cognitive flexibility required to process complex scenes. Current theoretical frameworks for attention are based almost exclusively on studies in visual cortex, where attentional effects are typically modest and excitatory. In contrast, attentional effects in auditory cortex can be large and suppressive. A theoretical framework for explaining attentional effects in auditory cortex is lacking, preventing a broader understanding of cortical mechanisms underlying attention. Here, we present a cortical network model of attention in primary auditory cortex (A1). A key mechanism in our network is attentional inhibitory modulation (AIM) of cortical inhibitory neurons. In this mechanism, top-down inhibitory neurons disinhibit bottom-up cortical circuits, a prominent circuit motif observed in sensory cortex. Our results reveal that the same underlying mechanisms in the AIM network can explain diverse attentional effects on both spatial and frequency tuning in A1. We find that a dominant effect of disinhibition on cortical tuning is suppressive, consistent with experimental observations. Functionally, the AIM network may play a key role in solving the cocktail party problem. We demonstrate how attention can guide the AIM network to monitor an acoustic scene, select a specific target, or switch to a different target, providing flexible outputs for solving the cocktail party problem.

## Author summary

Selective attention plays a key role in how we navigate our everyday lives. For example, at a cocktail party, we can attend to friend's speech amidst other speakers, music, and background noise. In stark contrast, hundreds of millions of people with hearing impairment and other disorders find such environments overwhelming and debilitating. Understanding the mechanisms underlying selective attention may lead to breakthroughs in improving the quality of life for those negatively affected. Here, we propose a mechanistic network model of attention in primary auditory cortex based on attentional inhibitory modulation (AIM). In the AIM model, attention targets specific cortical inhibitory neurons, which then modulate local cortical circuits to emphasize a particular feature of

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

sounds and suppress competing features. We show that the AIM model can account for experimental observations across different species and stimulus domains. We also demonstrate that the same mechanisms can enable listeners to flexibly switch between attending to specific target sounds and monitoring the environment in complex acoustic scenes, such as a cocktail party. The AIM network provides a theoretical framework which can work in tandem with new experiments to help unravel cortical circuits underlying attention.

## Introduction

A hallmark of cortical processing is the capacity for generating flexible behaviors in a context-dependent manner. A striking example of a problem that requires such cognitive flexibility is the cocktail party problem, where a listener can selectively listen to a speaker amongst other speakers [1]. Listening in such settings can be highly flexible, depending on the goal of the listener. For example, a listener can *monitor* the entire auditory scene, *select* a particular target, or *switch* to another target. Recent theoretical and experimental studies have begun to propose model networks and cortical mechanisms for producing flexible behaviors [2–8], and top-down control of cortical circuits via attention is thought to be a critical component.

The influence of attention on cortical processing has been intensively investigated in vision, resulting in a prominent theoretical framework of attention [9,10]. In contrast, relatively little is known about attentional mechanisms in auditory cortex. After the early discovery of “attention units” in the auditory cortex [11], there has recently been renewed interest on attentional effects in auditory cortex [12–14]. In comparison to the effects of attention in primary visual cortex, which are relatively small and excitatory [15], attentional effects in primary auditory cortex (A1) can be much larger and suppressive [13,16,17]. However, a theoretical framework for cortical mechanisms underlying auditory attention is lacking.

The responses of neurons in A1 can change rapidly when an animal is actively engaged in a task [8,13,16,17]. For example, cortical neurons with broad spatial tuning curves can sharpen tuning during attentive behavior [16]; whereas the spectral temporal receptive fields (STRFs) of cortical neurons with narrow frequency tuning can display the emergence of entirely new excitatory regions [17] or suppressive effects [13]. Cortical network mechanisms underlying such diverse attentional changes in tuning remain poorly understood. Changes in cortical tuning can also be driven by competing auditory stimuli in cocktail-party settings, even when an animal is anesthetized [18,19], suggesting the involvement of both bottom-up and top-down mechanisms [1,12,20,21].

There is a growing literature on computational models of auditory attention in the context of auditory scene analysis, as discussed in a comprehensive review [22]. These models can be grouped into bottom-up or top-down models. Bottom-up models have employed time-frequency representations of sound as an “auditory image” to compute salience maps using static or temporally evolving features of the image; and predictive coding theory to account for behavioral results in humans and animals processing auditory scenes [23,24]. Top-down models have formulated neural processing as spectral-temporal “filters” which extract features from the auditory image. In these models, attention adjusts the filter characteristics to optimize the detection and discrimination of targets to explain changes in receptive field properties in behaving animals [25]. Subsequent models have extended the feature analysis framework to propose computational principles, e.g., temporal coherence for linking multiple features across time (“streaming”) [26], and incorporated task structure to demonstrate that changes in

receptive field properties during behavior can be specific to task demands [27] as observed experimentally [28]. These previous computational models are all formulated in terms of statistical, signal processing or optimization principles. Thus, a key gap in current state-of-the-art models remains in formulating mechanistic models of *how* cortical computations are implemented by underlying cortical circuits. Indeed, as the review points out, “The field is particularly challenged by the lack of theories that integrate our knowledge of cortical circuitry in the auditory pathway with adaptive and cognitive processes that shape behavior and perception of complex acoustic scenes” [22]. So far, circuit-level models of cortical processing underlying the cocktail party problem have largely focused on bottom-up mechanisms [29,30]. Specific cortical circuit mechanisms underlying top-down attentional changes in cortical responses, and their functional role in solving the cocktail party problem, remain unclear [22].

Here we propose a network model to explain how experimentally observed cortical response properties in A1 could arise from underlying network mechanisms, via the interplay between bottom-up and top-down processes. Central to our network model is attentional inhibitory modulation (AIM), i.e., attention-driven modulation of distinct populations of cortical inhibitory neurons. Specifically, this mechanism relies on disinhibition of bottom-up cortical circuits, mediated via top-down inhibitory neurons, a prominent motif observed in cortex [5–8,31]. We first use the AIM network to model attentional changes in spatial and spectral tuning in auditory cortex [16,17], and then illustrate its potential functional role in solving the cocktail party problem.

## Results

### The AIM network

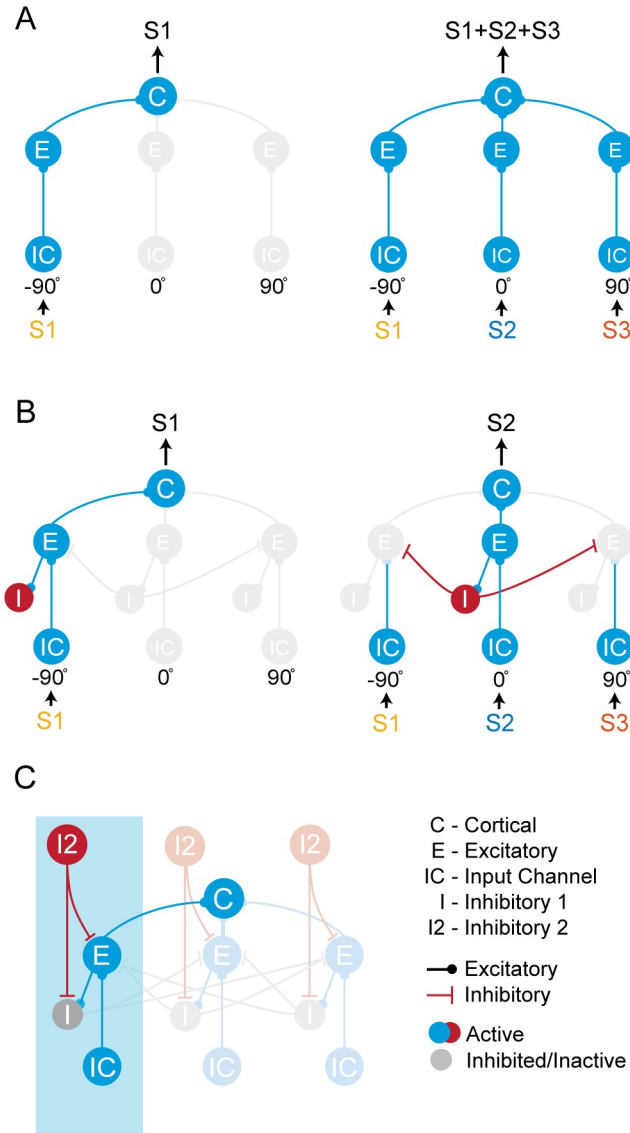
We began by focusing on spatial processing of multiple sound sources in auditory cortex, extending previous models of bottom-up processing (Fig 1A and 1B). The bottom-up network is made up of integrate-and-fire neurons and implements two key operations—integration and competition. Integration is mediated by broad convergence across spatial channels on the cortical neuron (C, Fig 1A), whereas competition is mediated by inhibition across spatial channels via I neurons (Fig 1B). These bottom-up mechanisms explain two key features observed experimentally in anesthetized or passive animals: broad spatial tuning of cortical neurons to single sounds, and sharpening of spatial tuning in the presence of multiple competing sounds [18,19,29,30].

We extended the bottom-up network to model the effects of attention in the AIM network (Fig 1C). Previous studies have shown that bottom-up cortical representations can be modulated in the attentive state by distinct sub-types of inhibitory neurons. To model such attentional inhibitory modulation, we introduced an additional layer of inhibitory neurons (I2, Fig 1C). I2 neurons can control the spatial and spectral tuning of the cortical neuron in different attentive states, by modulating the activity of E and I neurons in the bottom-up network.

### Attentional changes in spatial tuning

A previous experimental study in cat A1 demonstrated that the spatial tuning of cortical neurons sharpen during attentive behavior [16]. Specifically, the authors observed that A1 neurons exhibited broad spatial tuning when the animal was idle but sharpened their spatial tuning when the animal performed a spatial localization task, as demonstrated by the changes in the azimuth-dependent peristimulus time histograms (PSTHs) (Fig 2, 3<sup>rd</sup> column). We modeled this attention-induced sharpening effect using the AIM network.

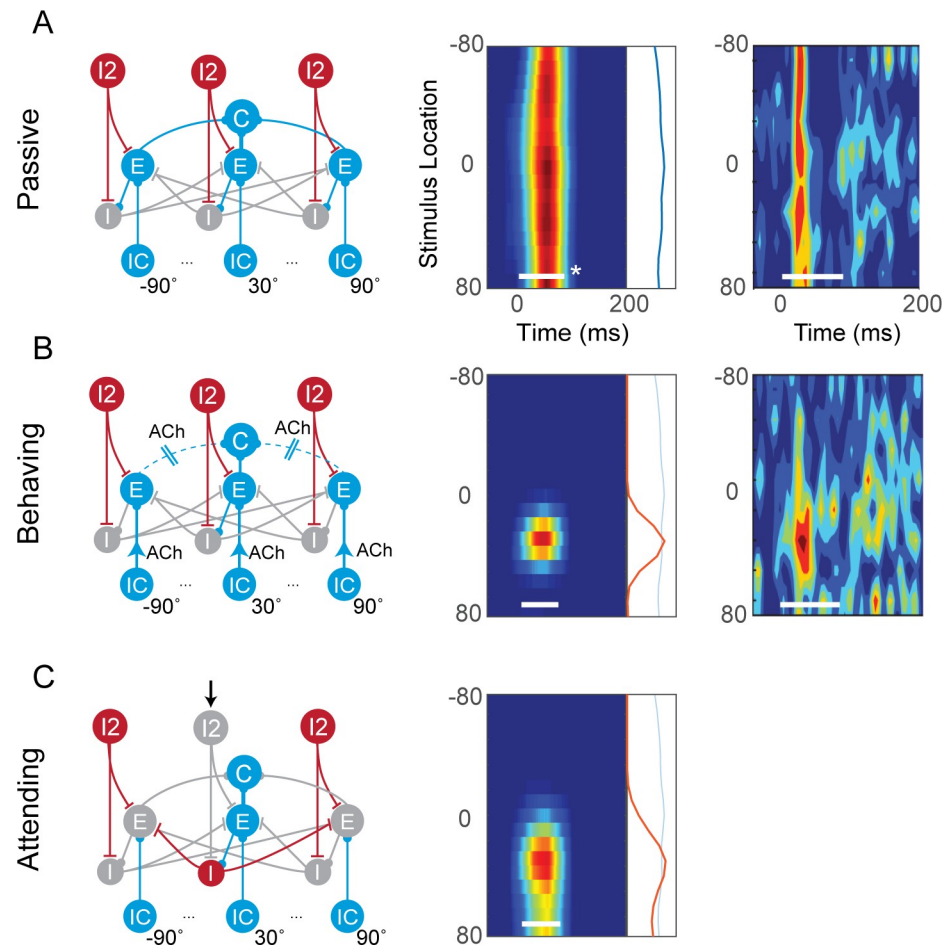
In this simulation, the AIM network consisted of an array of spatial channels tuned to locations between  $-90^\circ$  and  $90^\circ$  azimuth. We then probed the network with broadband noise from



**Fig 1. Sub-networks within the AIM network.** (A) A convergence network. S1, S2, and S3 denote distinct audio stimuli 1, 2, and 3, which are placed virtually at  $-90^\circ$ ,  $0^\circ$ , and  $90^\circ$  azimuth, respectively. (B) A passive switching network realized with the addition of I neurons. (C) The AIM network, realized with the addition of I2 neurons, which modulates the I neurons. One spatial channel is highlighted in blue. Neurons in the same spatial channel process the stimuli from a specific spatial location. Response of the cortical neuron, C, represents the output of the network.

<https://doi.org/10.1371/journal.pcbi.1009356.g001>

various spatial locations and analyzed the C neuron’s response as a function of space. Based on the azimuth-dependent PSTHs, we found that when I2 neurons in all spatial channels were on, the cortical neuron in the AIM network exhibited broad tuning—similar to the idle condition in the experiment (Fig 2A). The release of acetylcholine (ACh) during behavioral task performance can suppress intracortical excitatory connections and strengthen thalamocortical connections [32–34]. When we simulated these effects in the AIM network, the spatial tuning of the cortical neuron sharpened, resembling the tuning in the behaving condition in the Lee and Middlebrooks study (Fig 2B). In that study, however, animals were not required to attend to a specific location during the task. We simulated selective attention to a specific location by



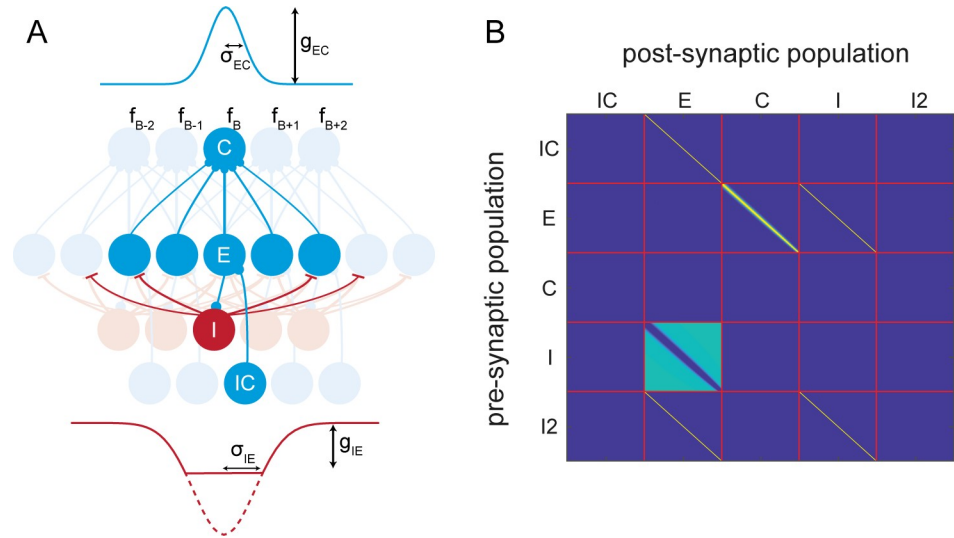
**Fig 2. Attentional sharpening of spatial tuning.** First column shows the spatial AIM network in (A) passive condition; (B) Behaving condition, where the network is modulated with the cholinergic system; (C) Attending condition, where selective attention is simulated by controlling the state of specific I2 neurons. Second column shows the spatial tuning, i.e., PSTH expanded vertically to show the spatial dimension, calculated from the network output. Firing rate is shown (red is higher) as a function of stimulus location over time, in response to a broad band noise. Third column shows experimentally recorded spatial tuning in the cat A1. \*White bars indicate the duration of the noise stimulus (80ms). Figures from Lee and Middlebrooks reproduced here with the permission of Nature Neuroscience.

<https://doi.org/10.1371/journal.pcbi.1009356.g002>

inactivating an I2 neuron in a specific channel, e.g.,  $30^\circ$  (Fig 2C, left). We found that, in this case, the spatial tuning of the cortical neuron also sharpened (Fig 2C, right). In the AIM network, this effect occurs because of two key mechanisms: the disinhibition of the attended channel by the I2 neuron, which then drives powerful inhibition of competing channels by the I neuron. Thus, selective attention activates focal disinhibition at the attended location and suppression at other locations in the network.

### Attentional changes in spectral tuning

Rapid changes in receptive fields during task performance, thought to arise from attentional mechanisms, have also been observed in the frequency domain as in the experiments by Fritz et al., (2003) and Atiani et al., (2009). Here, we show that the attentional mechanisms in the AIM network can also account for these experimental observations.



**Fig 3. AIM Network Diagram for the spectral network.** A) convergence of connectivity across neuron types. E neurons converge to C neurons locally, centered around a neuron at the best frequency,  $f_B$ . The connectivity strength decays across adjacent channels, and is modeled by a Gaussian function. The convergence width is determined by  $\sigma_{EC}$ , and the connectivity strength is determined by  $g_{EC}$  (see Methods). The spread of inhibitory connection from I neurons to E neurons is modeled by a thresholded, inverted Gaussian function. The connectivity within  $2\sigma_{IE}$  of this Gaussian function is zero. The inhibition gradually rises to a value determined by  $g_{IE}$ . B) The network connectivity matrix which describes all connections within the network for these simulations.

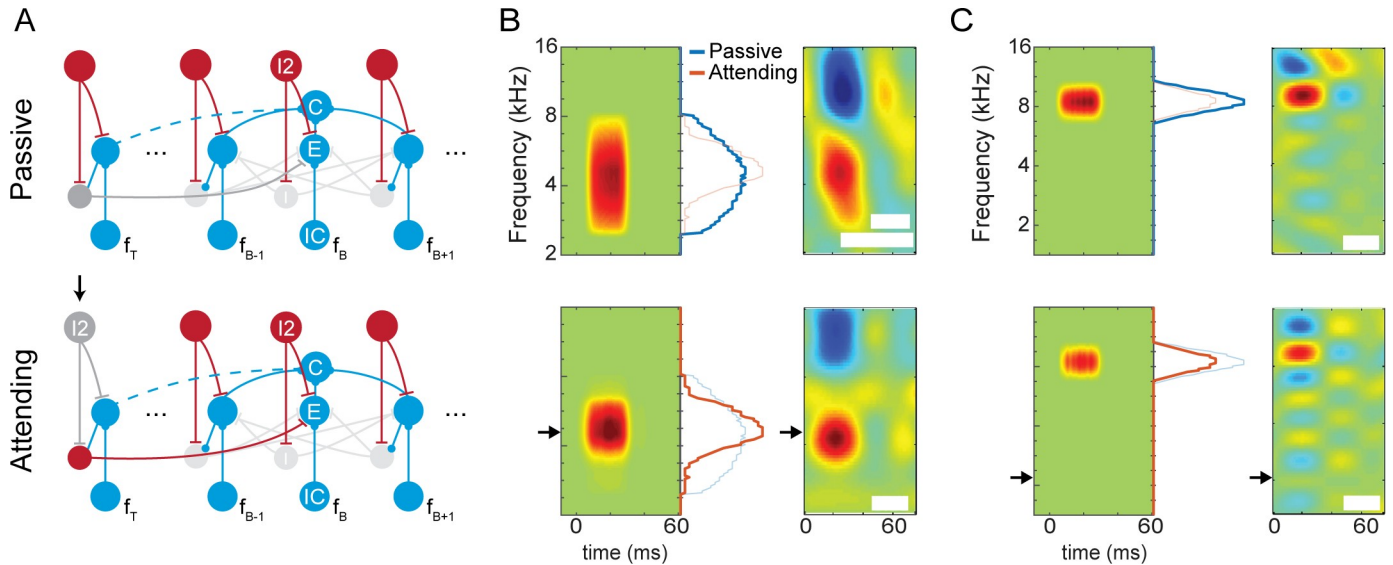
<https://doi.org/10.1371/journal.pcbi.1009356.g003>

We first constrained the AIM network to incorporate experimentally observed features of network connectivity in the frequency domain (Fig 3). Unlike the broad connectivity in the spatial domain (Figs 1 and 2), connectivity across frequency channels is localized to nearby frequencies, reflecting the tonotopic organization of the auditory cortex.

We then simulated the spectral tuning of neurons in the network in the passive vs. attentive states. Here, we probed the network with pure-tone stimuli in the frequency ranges shown in Fig 4. In the attentive state, the network attended to a specific target frequency, distinct from the best frequency of the neuron, as in the experiments by Atiani et al. [35]. Atiani et al. showed the neuron’s response as spectrotemporal receptive fields (STRFs). Here, we used frequency-dependent PSTHs as an approximation to STRFs in order to compare our results.

In the attentive state, the I2 neuron in the attended channel (*i.e.*, the target frequency channel,  $f_T$ , Fig 4) is suppressed during attention, disinhibiting the E and I neurons in that channel (Fig 4A). In this case, we found that when the target frequency was close to the best frequency of the neuron, attention produced an increase as well as a sharpening of the response near the best frequency, as observed experimentally (Fig 4B). In the AIM network, this occurs because when the target frequency is close to the best frequency, excitation from the disinhibited E neuron in the target channel increases the peak response of the cortical neuron, and inhibition via the I neuron in the target channel sharpens the shape of response. In contrast, when the target frequency is far from the best frequency of the neuron, the effect due to inhibition dominates, producing a net suppressive effect on the response, which is also observed experimentally (Fig 4C). Thus, the AIM network qualitatively explains salient features of the experimental observations by Atiani et al.

In addition to sharpening, strengthening, and weakening of STRF hotspots, the experimental results of Fritz et al. showed that a secondary hotspot in the STRF may arise when the animal is in the attentive state. (Fig 5).



**Fig 4. Simulating attentional effects in Atiani et al.** A) The AIM network in passive and attending state. Black arrow marks the activation of attention on the target frequency channel. The target frequency  $f_T$  represents the attended frequency, while  $f_B$  represents the neuron's best frequency. Dashed blue line indicates that connections to frequency channels far from  $f_B$  are weaker than connections to frequency channels close to  $f_B$ . B) The results of the AIM network (left panels) versus the results described in Atiani et al. (right panels), when  $f_T$  (black arrow) is near  $f_B$ . Marginals show the total spiking activity across the duration of the simulation. C) The results of the AIM network versus the results shown by Atiani et al., when  $f_T$  is far from  $f_B$ . Figures from Atiani et al. are reproduced here with the permission of Neuron. White patches hide irrelevant text.

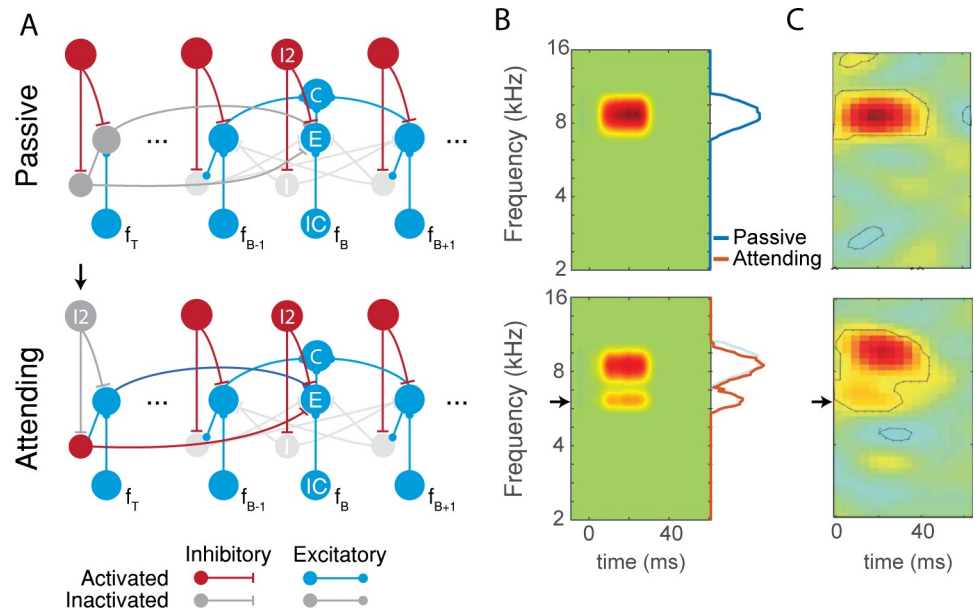
<https://doi.org/10.1371/journal.pcbi.1009356.g004>

We hypothesized that this is a result of the strengthening of an intracortical connection between the target frequency and the best frequency in the attentive state (see *Discussion* for possible mechanisms). To test this hypothesis, we added an additional E-E connection between the  $f_T$  and  $f_B$ , and found that in the passive state, the neuron responded to frequencies near its best frequency, showing a single hotspot. On the other hand, in the attentive state, the same neuron also responded to the target frequency, as seen by the emergence of a new excitatory region at the target frequency in Fig 5B. This effect occurs due to the strengthening of synaptic connection between the target frequency and the best frequency in the attentive state. There is also a suppressive effect on the response to the best frequency, as seen in the slight reduction in amplitude of the tuning curve at best frequency. This effect occurs due to inhibition driven by the I neuron in the target frequency channel. Both of these effects were observed in experimentally measured receptive fields [17]. Thus, the AIM network qualitatively explained salient features of experimentally observed changes by Fritz et al.

### Functional implications

We hypothesized that the attentional mechanisms in the AIM network play an important functional role in processing complex auditory scenes. Two highly effective mechanisms for sound segregation are spatial hearing and frequency selectivity. We first considered functional implications in the spatial domain.

When entering a cocktail party, one might want to monitor the entire scene, focus one's attention on a conversation partner, or switch attention between conversation partners. How does top-down attentional control modulate bottom-up mechanisms to enable this flexible behavior? To illustrate the behavior of the AIM network in these different modes, we simulated several scenarios. In these simulations, we used three spatial channels corresponding 0°, 90°, and 180°.



**Fig 5.** Simulating the effects in Fritz et al., A) The proposed mechanism underlying the changes in A1 behavior. Frequency channels adjacent to the best frequency  $f_B$  as well as the target frequency  $f_T$  are shown. Black arrow indicates the target of selective attention. The dark blue connection highlights the additional intracortical connection unique to this simulation. B) Simulated A1 neuron STRFs in the two states of attention. The model qualitatively reflects the changes observed in physiological recordings. C) Physiological recording of a STRF of an A1 neuron when the animal is passively listening (top) and when the animal is attending to a target tone, marked by the black arrow (bottom). Figures from Fritz et al. are reproduced here with the permission of Nature Neuroscience.

<https://doi.org/10.1371/journal.pcbi.1009356.g005>

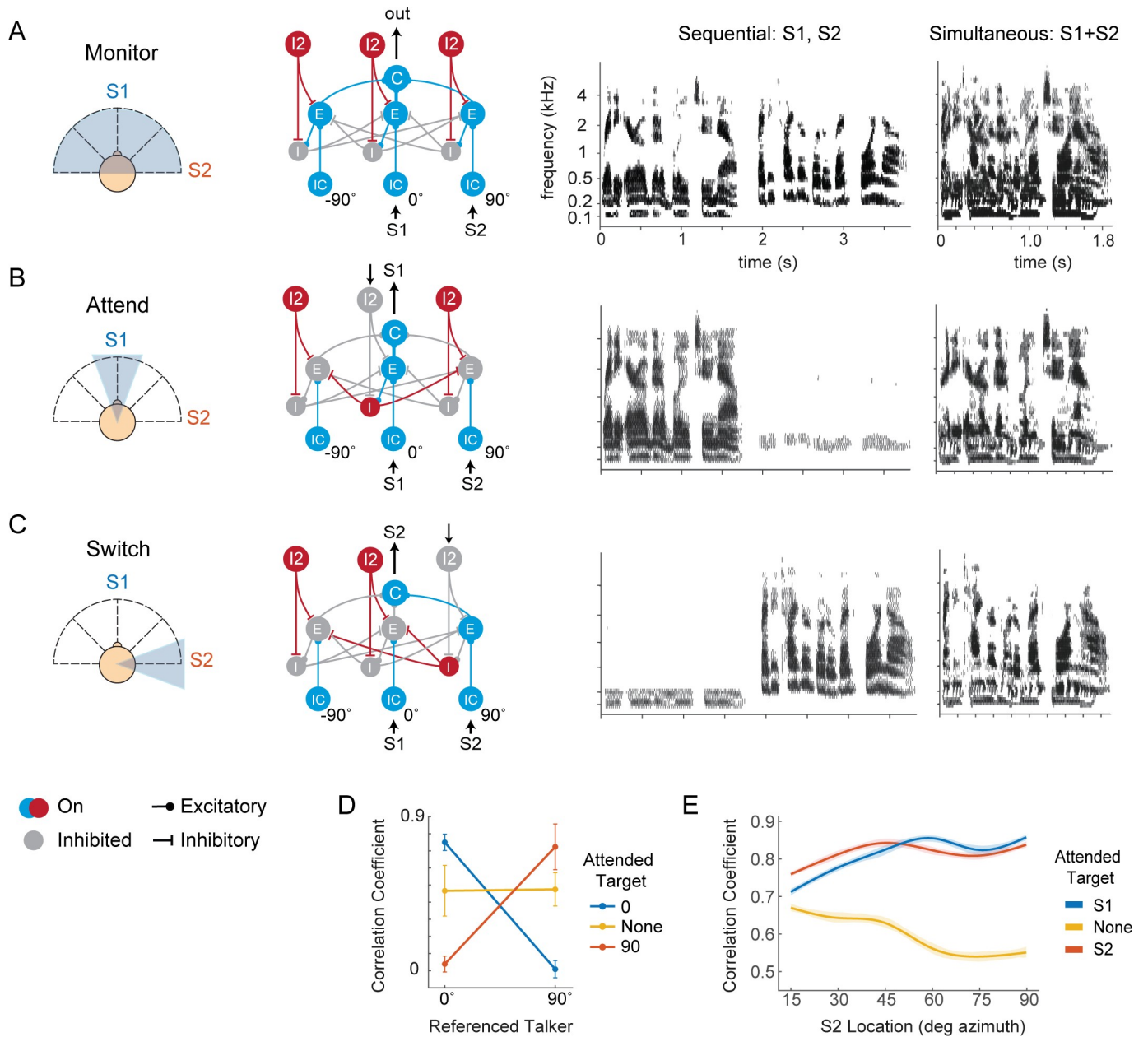
and  $-90^\circ$  azimuths, presenting the network with different tokens of speech stimuli at these locations, either sequentially or simultaneously (see [Methods](#)).

To demonstrate passive listening (the “monitor” mode), we set all I2 neurons active, thereby silencing all I neurons ([Fig 6A](#)). In this case, when the speech tokens were presented sequentially to the network, the network output resembles each individual speaker ([Fig 6A](#)). When the speech tokens were presented simultaneously, the network output resembles their mixture. Thus, in this mode, the network broadly monitors the acoustic scene across different spatial locations.

To simulate selective attention to a particular speech token, we first inactivated the I2 neuron in the  $0^\circ$  spatial channel, thereby activating lateral inhibition via the disinhibition of I neurons in that channel ([Fig 6B](#)). In this case, the network output resembled the  $0^\circ$  speaker output, regardless of whether the speakers were presented sequentially or simultaneously. Finally, to simulate switching attention to a different location ( $90^\circ$ ), we turned off the I2 neuron at  $90^\circ$ , activating lateral inhibition via the disinhibition of I neuron at that location ([Fig 6C](#)). In this case, the network output was more similar to the output for the speaker at  $90^\circ$  ([Fig 6D](#)).

Finally, we demonstrate the effect of speaker separation on network performance. First, we expanded the network to have seven spatial channels tuned from  $0^\circ$  to  $90^\circ$  azimuth in  $15^\circ$  increments. We then kept the first speaker at  $0^\circ$  while varying the location of the second speaker, and presented both speakers to the network simultaneously. We found that the network output became more representative of the attended targets (either speaker 1 or speaker 2) as the two speakers separate in space, demonstrating the effect of spatial release from masking [36] ([Fig 6E](#)). In summary, when an I2 neuron in a specific spatial channel is inactivated, it disinhibits the I neuron at that location, causing the network to selectively attend to that spatial



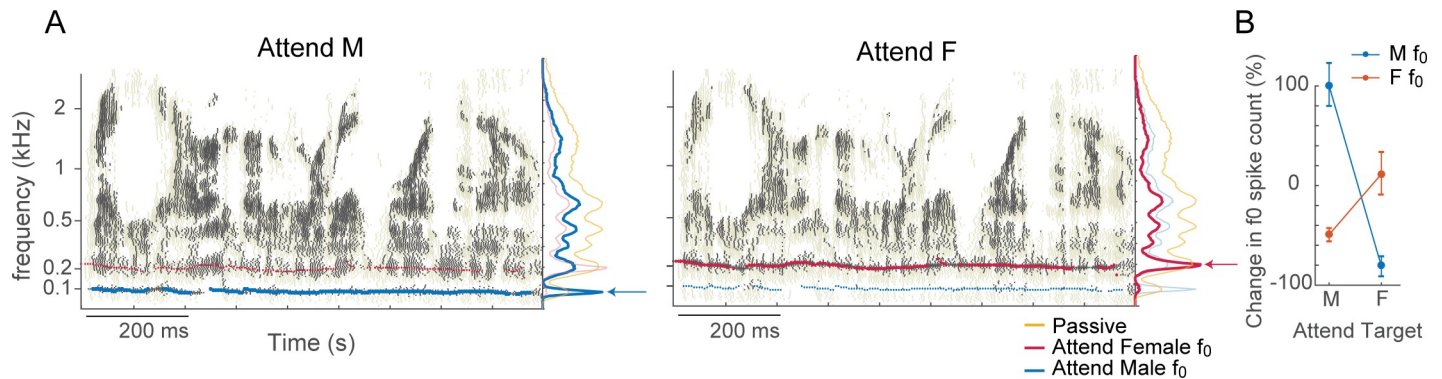


**Fig 6. Functional implications of the AIM network: Spatial tuning of the network is dictated by the state of TD neurons.** (A) The network monitors the entire azimuthal plane when all TD neurons are active. (B) The network attends to a specific direction if the corresponding TD neuron is off. (C) The network attends to a different location if a different TD neuron becomes inactive. Column 3 shows the result of simulations when speakers are presented sequentially to the network, in spike rasters. Column 4 shows the result of simulations when speakers are presented simultaneously to the network. (D) The AIM network can recover an attended target within a speech mixture, as quantified by the cross-correlation measures between the simultaneous simulation network output and single speaker spike rasters. Error bars show standard deviation ( $n = 20$ ). X-axis is the reference speaker, and each line color denotes the attended location. (E) Spatial separation of two talkers (S1, S2) vs. network performance, as quantified by correlation between the network output to the attended target. The encoding becomes more representative of the attended speaker as the separation between the two speakers increases. In the “not attending” case, S1 is used as the reference for correlation calculation. Shaded area represent 95% confidence interval,  $n = 20$ .

<https://doi.org/10.1371/journal.pcbi.1009356.g006>

location. Additionally, the ability of the network to attend to a target depends on the spatial separation between the target and its maskers.

We next investigated the functional role of the AIM network in the frequency domain using the same network shown in Fig 4. In this simulation, two competing speech tokens, a



**Fig 7. Functional role of the AIM network in the frequency domain.** (A) AIM network outputs (raster plots) when attending to the male talker's  $f_0$  (left) or female talker's  $f_0$  (right). Yellow rasters in the background shows the network output in the passive condition. Blue and red lines mark the estimate  $f_0$  for each speaker. Blue, red, and yellow lines in the marginal show the total spike counts per frequency channel for attend male, attend female and passive conditions. (B). Change in spike count (%) in the male or female  $f_0$  channel when attending to the male or female target, compared to the passive condition. Error bars show standard deviation ( $n = 20$ ).

<https://doi.org/10.1371/journal.pcbi.1009356.g007>

male and female speaker, originate from the same spatial location. In this case, differences in spatial cues cannot be exploited for segregation, but differences in spectral features of the two speakers, e.g., the fundamental frequency (F0), are available. We simulated the network in three conditions: passive, attending to male F0, or attending to female F0. Attention was simulated (as in Fig 2) by inactivating the I2 neuron which disinhibits the I neuron at the attended frequency channel. In the attending modes, network activity showed a sharp peak around the F0 of the attended speaker, and a suppression of activity around the F0 of the competing speaker (Fig 7). Similar to the earlier results shown in Fig 3, the attended F0 received an increase in spiking activity in that region while spiking activity in frequencies far from the F0 are suppressed (Fig 7A marginals and Fig 7B).

## Discussion

The capacity for generating flexible behaviors in a context-dependent manner is central to many complex cognitive tasks. How cortical circuits achieve such flexible computations is a central area of investigation in both theoretical and experimental neuroscience. Recent theoretical studies have begun to propose model networks capable of producing flexible behaviors, e.g., gating mechanisms for flexible routing of information [2–4], and experimental studies have begun to reveal cortical mechanisms underlying flexible gating of information and attentional control [4,6,37]. However, such models are lacking for auditory cortex. In this study, we propose the AIM network, which describes a mechanism of interaction between top-down and bottom-up processes in auditory cortex that may underlie the attention-driven changes in cognitive behavior.

### Flexible cortical processing

The rapid flexibility of the AIM network is generated by top-down inputs, which control the state of the network by dictating the on/off state of specific I2 neurons. The top-down inhibition of I2 neurons disinhibits the I neuron of the attended channel, which then suppresses competing channels via top-down lateral inhibition, resulting in focused attention. On the other hand, when I2 neurons in all channels are active, the network integrates information from all input channels. In the spatial case, this produces broad spatial tuning and allows the network to monitor the entire scene. The ability to switch between these behaviors is important from a functional standpoint. A network that is always selective for a single channel may

fail to detect important events in the scene at other locations. The different states of the I2 layer (e.g., all on vs. one off), allows the opportunity for *exploration* (by detecting events across a broad range of locations) as well as *exploitation* (by selectively listening to a particular channel) in a dynamic manner.

Switching between exploration and exploitation is especially interesting in the context of findings that the “spotlight of attention” fluctuates in a rhythmic manner [38]. In our model, such fluctuations in the strength of the top-down input would cause the network to alternate between periods favoring broad detection of sounds across the entire acoustic scene and fine discrimination at a single spatial location [39]. Such periods may allow salient sounds in the background to capture the spotlight of attention, as demonstrated in a recent study in humans [40]. Moreover, changes in the location of the top-down input would promote switching attention to a different location [41]. These various behaviors may play important roles in how animals navigate complex environments.

### Alternative/additional mechanisms

The AIM network describes the mechanistic interaction between top-down and bottom-up processes, and even though it can explain various experimental observations, alternative/additional mechanisms may be involved in some aspects of the experiments modeled here.

**Neuromodulation.** The neuromodulatory systems can exert powerful control over the *global* state of cortical networks, e.g., asleep, quiet arousal, and active attention [42]. Two key modulatory projections to cortex involve norepinephrine (NE) and acetylcholine (ACh), which have been implicated in arousal and attention, respectively. In our network, we assumed that when the network is in the “monitor” mode, the I2 layer is on. Such a global state may correspond to arousal and be NE-dependent. Top-down suppression of I2 neurons in our model could correspond to an attentive state and be ACh-dependent.

Indeed, we showed that global cholinergic effects on cortical circuits, i.e., suppression of intracortical connections and enhancement of thalamocortical connections, could also produce a sharpening in spatial tuning (Fig 2). In the study by Lee and Middlebrooks [16], animals were not required to attend to a fixed location when performing the localization task. Thus, sharpening of attention via the activation of global cholinergic mechanisms may be more consistent with that experimental design.

**Top-down inputs from frontal areas.** Although cholinergic mechanisms are clearly important in attentional states, such mechanisms are thought to operate on slow timescales and can be long lasting [42], whereas switching between exploration and focused attention requires rapid, reversible changes in cortical outputs, potentially on sub-second timescales. Recent studies with EEG and fMRI in humans have suggested top-down activation in the frontal areas modulates processing in auditory cortex on the time scale of hundreds of milliseconds [43,44]. Evidence from studies in mice also support the idea that the frontoparietal network can modulate processing in the primary sensory cortices during selective attention [5,45–47]. Together, these studies suggest that the top-down signals responsible for modulating the AIM network may originate from the frontal areas.

**Synapse-specific gating.** Our model predicts that experiments where animals are required to selectively attend to a specific location should also produce a sharpening of spatial tuning in A1. For the experiments by Fritz et al. [17], we found that strengthening of the intra-cortical synaptic connection between the target frequency and the best frequency could explain the emergence of new excitatory regions at the target frequency. A possible mechanism for transient, reversible strengthening of intracortical synapses is synapse-specific gating [3,4], which may then promote long-term strengthening via classic Hebbian plasticity.

## Cortical inhibitory neurons and function

Inhibitory neurons play key roles in the AIM network. There are several types of inhibitory neurons in cortex [48]. The majority of inhibitory neurons can be placed into three categories: those that express parvalbumin (PV), somatostatin (SOM), or vasointestinal peptide (VIP). It is worth noting that most currently available information on specific classes of interneurons come from studies in rodents in a variety of cortical areas, whereas key experimental observations modeled in this study were obtained in other species. Thus, it is difficult to directly map the functional groups of neurons, e.g., I2 and I, in the model to identified interneuron types e.g., PV, SOM and VIP neurons. Nevertheless, we suggest some hypotheses on a possible correspondence based on recent experiments in rodents, to motivate future experimental work.

**VIP neurons.** The top-down input in the model could correspond to inputs from VIP neurons. VIP mediated inhibition is engaged under specific behavioral conditions, including attention [5,7]. It has been proposed that VIP cells “open holes in the blanket of inhibition” [49], generating the “spotlight of attention” [5]. Our results are consistent with this intuition, with the top-down input being critical for selecting a particular target and switching to a different target.

VIP input is often thought to favor excitation, due to the disinhibition of excitatory neurons [50]. In the model, top-down inhibition of an I2 neuron in a specific channel activates powerful inhibition via I neurons that suppress competing channels, leading to the selection of the target. Thus, the model also explains powerful suppressive effects of selective attention, which have been observed in auditory cortex [13]. The model predicts that silencing top-down inputs to a specific channel, via optogenetics or other methods, should block the effects of selective attention.

**SOM and PV neurons.** VIP neurons are known to inhibit SOM neurons, which in turn inhibit excitatory neurons [31]. This motif suggests that the I2 neurons in our model may correspond to SOM neurons, specifically Martinotti cells, which are strongly targeted by VIP neurons [51].

The I neurons mediate powerful and sustained inhibition of competing channels in the model. A key distinction between this type of inhibition and “classical” lateral inhibition observed at multiple stages of sensory processing starting at the periphery is noteworthy. Classical lateral inhibition is activated by bottom-up stimulus-driven mechanisms, whereas the inhibition in our model is driven by top-down attentional mechanisms. To distinguish these two cases, we refer to the inhibition mediated by I neurons in the model as “top-down lateral inhibition”. This distinction is conceptually and functionally important, because unlike classical lateral inhibition, which is recruited automatically by the stimulus, top-down lateral inhibition can be recruited volitionally. Such top-down lateral inhibition can be activated by direct disinhibition of I neurons, disinhibition of E neurons which drives feedback lateral inhibition via I neurons (S1 Fig), or a combination of the two.

In principle, top-down lateral inhibition could be mediated by any interneuron type. Although PV neurons are a possible candidate, the long-lasting inhibition required to suppress competing channels in our model should be distinguished from the fast and transient dynamics of inhibition typically associated with PV neurons [48]. SOM neurons can also mediate feedback lateral inhibition to generate a “winner-take-all” circuit and suppress competing channels [52], or modulate bottom-up inputs in specific layers [51,53]. Developing behavioral paradigms for investigating attention in rodents combined with optogenetic manipulations, and/or developing methods for selectively manipulating different interneuron types in other species, are promising future directions for identifying specific cell-types involved in mediating top-down lateral inhibition.

## Space vs. frequency

We related the effects of attention in the AIM network to key experimental observations on changes in cortical spatial and frequency tuning in animals engaged in a behavioral task vs. passive animals [16,17]. Similar changes have also been reported in the primary auditory cortex of humans [54]. Assuming that attentional mechanisms are a key factor in driving such changes [12,16,55], the effects of attention appear very different in the spatial and frequency domains. In the spatial domain, broad tuning sharpens during task performance, whereas in the frequency domain, narrow tuning can be enhanced or suppressed depending on the target frequency and other parameters such as SNR. Our results suggest these apparent differences in the spatial vs. the frequency domain may share similar underlying attentional mechanisms.

A key difference between the spatial and frequency domains in our model is the convergence from the E neurons to the C neuron, which is broad in the spatial domain but narrow in the frequency domain. Previous studies in the auditory cortex have found a tonotopic organization, but no topographic organization for spatial tuning [56]. Therefore, local synaptic connections in a patch of cortex may result in convergence from neurons with similar tuning in frequency but a broad range of tuning in space, which is consistent with our model. Thus, our results suggest that the same cortical mechanisms underlying attention can produce diverse effects on stimulus tuning, due to differences in the cortical organization of stimulus features, e.g., space or frequency. For simplicity, we considered spatial and spectral networks separately. Future models should unify these two dimensions.

## Domain-specific considerations

**Spatial domain.** Here, we did not explicitly model how spatially tuned inputs to the AIM network arise, an aspect that is likely to be species dependent. In the AIM network, spatial tuning is inherited from tuning for acoustic cues in pre-cortical areas [30], perhaps the simplest scenario consistent with experimental observations [57–59]. Additional mechanisms, e.g., forward suppression, may further sharpen or generate spatial tuning in cortex [60,61]. In rodents, spatially tuned responses covering a range of azimuths have been observed in cortical areas [62], and may emerge from excitatory-inhibitory interactions in the underlying network [63]. From a functional standpoint, it is interesting to note that sharp tuning is not necessary for the monitor mode, but only for the selective mode of the AIM network. In some species, the sharpness of tuning may emerge in the attentive state based on state-dependent mechanisms, and/or inputs from other brain areas, e.g., the superior colliculus, which shows a map of auditory space [64,65] and can modulate responses in A1 via the pulvinar [66]. These outstanding issues will require further experimental work, especially in attentive animals, as well as the development of species-specific models.

**Spectral domain.** The model has several simplifications and limitations that motivate future directions of work. For example, in the frequency network, we used pure tones to characterize the responses of neurons in the network and relate them to experimentally observed STRFs obtained with ripple noise stimuli. Although this approach captured salient attentional effects observed experimentally, future studies should probe non-linear components using complex stimuli, e.g., ripples and natural sounds.

In this study, we focused on excitatory regions of STRFs, modeling three representative changes in the excitatory regions of STRFs observed by Fritz et al and Atiani et al. [35]. However, it is known that the balance between both excitatory and inhibitory subregions can play an important functional role [67]. In preliminary simulations we found that attention could also decrease an inhibitory subregion, as observed by Fritz et al. (S2 Fig). Modeling the effects of attention on inhibitory regions, and complex STRFs with both excitatory and inhibitory

regions, merit further investigation in the future. For example, the diverse effects of training paradigms on inhibitory regions [28], will require modeling the effect of training, reward or punishment on cortical circuits in the model.

We used the AIM network to illustrate how top-down inhibition of I2 neurons can enhance the representation of sounds at an attended location, or an attended feature, e.g., F0 in speech, by suppressing competing sounds at a different location or F0. F0 is likely to be one of many potential features that contribute to speech segregation. However, a similar principle could be applied to more complex features, e.g., enhancing the representation of an attended harmonic, employing harmonic template neurons in the auditory cortex [68]. Such enhancement of an attended feature accompanied by suppression of competing features may contribute to speech segregation in settings where spatial cues are unavailable [69].

**Temporal domain.** Temporal dynamics likely play a large role in auditory processing. For example, the neurons observed by Lee and Middlebrooks are highly sensitive to stimulus onset (Fig 2). Our work focused on the effect of attention in the spatial and spectral domain, and thus we did not include detailed models of temporal dynamics, e.g., adaptation or sensitivity to stimulus onset or offset, or investigate the transient tuning properties during attention switching. Additionally, we did not investigate temporal phenomena that is likely to play an important role in speech segregation, e.g., temporal aspects of F0 or tracking of slow spectrotemporal modulations [70], and auditory “streaming” for linking sound segments over time [71]. We believe that thoroughly investigating these aspects at the cortical circuit level will require modeling rich temporal aspects of neuronal and network dynamics, e.g., adaptation, synaptic facilitation and depression, oscillations, synchrony, and coherence, and is outside the scope of this current study. Future extensions of the AIM network should incorporate these aspects to link mechanisms of neuronal and network dynamics to attentional dynamics.

## A mechanistic model of attention

Previous studies have modeled the effects of attention on auditory cortical receptive fields using mathematical and computational principles such as temporal coherence [22,72]. In contrast, the AIM network is a cortical circuit level model underlying attentional effects. One recent study modeled different STRFs in the attending vs passive state of the ferret A1 with a two-layer spiking network [73]. The focus of that study was to produce detailed fits of STRFs in attending and passive animals. In contrast, the focus of this study was to propose general cortical circuit mechanisms, e.g., top-down disinhibition, underlying the effects of attention on both spatial and spectral tuning. Another previous study modeled global cholinergic mechanisms underlying changes in STRF [74], similar to the effects modeled in Fig 2B. However, that study did not include the selective top-down disinhibitory mechanism, which was unknown at that time and is a key mechanism in the AIM network.

Original models of attention in vision were also developed based on computational principles, e.g., biased competition or normalization [9,10]. At the time, available information on cortical circuits to guide and constrain circuit-based models were limited. Subsequently, cortical circuit-based models of visual attention have been proposed [75]. With the rapidly emerging knowledge of specific cell types and circuitry in auditory cortex, along with the availability of powerful optogenetic tools for cell type-specific perturbations, the AIM network may help guide the design of new experiments to unravel cortical circuits that underlie general attention.

## Methods

Simulations and models were implemented in Matlab (Natick, MA, USA).

## Stimuli

Three sets of auditory stimuli were used, depending on the specific simulation. White Gaussian noise was used as the stimulus in spatial tuning simulations, pure tones with frequencies approximately equal to the center frequencies of the gammatone filterbank (see Subcortical Processing) were used in spectral tuning simulations, and speech tokens from the Coordinated Response Measure (CRM) corpus were used in the functional demonstration simulations [76]. In spatial simulations where stimuli were placed along the azimuth, directionality is imparted on the stimuli by convolving them with the head-related transfer functions (HRTFs) of the Knowles Electronics Mannikin for Acoustic Research (KEMAR) [77,78].

## Subcortical processing

Stimuli for each simulation were first processed and encoded with models of the auditory periphery and midbrain, then presented to the network. The auditory periphery was modeled by a gammatone filterbank, implemented using the Auditory Toolbox [79]. It was used to separate the sentence mixture into 64 narrowband frequencies, with center frequencies ranging from 200 to 8000 Hz, uniformly spaced on the equivalent rectangular bandwidth scale.

We used a previously published model of the midbrain to perform spatial segregation of spatialized stimulus mixtures, as well as to encode the stimuli. If a simulation did not use spatial stimuli as the input, the stimuli were treated as dichotic. For details pertaining to the midbrain model, see Fischer et al. and Chou et al. [30,80]. Briefly, the midbrain model computed binaural features (i.e., interaural timing and level differences) in each time-frequency tile (i.e., narrowband and short time window). Model neurons encoded the stimulus at specific time-frequency tiles if the binaural features of the stimuli matched the “preferred” binaural features of the model neuron, thereby performing spatial segregation. The preferred binaural feature of each model neuron is specific to the frequency and spatial channel each neuron belonged to. There were 64 frequency channels in the midbrain model, corresponding to each channel of the gammatone filter. The number of spatial channels in the midbrain model depended on each specific simulation. The input neuron in a spatial channel is spatially tuned to the azimuth corresponding to that channel, consistent with spatial tuning of acoustic cues observed in subcortical areas [57–59]. The spiking responses of these model neurons were used as the input to the AIM network.

## Attentional inhibitory modulation (AIM) network

The AIM network was implemented using the DynaSim package [81], and its structure is illustrated in Fig 1. For simplification purposes, only one frequency channel and three spatial channels are shown. A “spatial channel” refers to the sub-network of neurons that are responsible for processing inputs from a specific spatial location (blue shading, Fig 1). The number of spatial and frequency channels in the network, and their connectivities, depended on the specific simulation being explored.

Five neural populations were created within the network: excitatory input (IC), excitatory (E), inhibitory (I), output cortical (C), and a second inhibitory (I2) population. IC neurons represent the bottom-up inputs to the network from the subcortical model. I2 neurons represent attentional top-down control. With the exception of the C neurons, a number of neurons were created within each population, corresponding to each of the spatial or frequency channels needed in a simulation. All five neural populations are implemented as leaky integrate-and-fire neurons whose dynamics are defined by the following differential

equation [82]:

$$\frac{dV}{dt} = \frac{(g_{leak}(E_{leak} - V) - i_{syn})}{C}$$

where  $V$  is the membrane potential,  $i_{syn}$  is the synaptic input current,  $C$  is membrane capacitance,  $g_{leak}$  is the membrane conductance, and  $E_{leak}$  is the equilibrium potential. The spike-and-reset mechanism employed in our model dictates that if  $V > V_{thresh}$ , then  $V \rightarrow V_{reset}$ . Here,  $V_{thresh}$  is the action potential threshold and  $V_{reset}$  is the reset voltage. Values for these parameters are listed in Table 1.

The dynamics of the synaptic input current is defined by a double exponential:

$$i_{syn}(t + 1) = i_{syn}(t) + g_{syn} \left[ \left( e^{-\frac{t}{\tau_D}} - e^{-\frac{t}{\tau_R}} \right) u(t) \right] (netcon) (V - E_{syn}) + i_{app}(netcon)$$

where  $t$  is time since the previous spike,  $g_{syn}$  is the synaptic conductance,  $\tau_D$  and  $\tau_R$  are the decay and rise time constants, respectively, and the difference of exponentials represent the excitatory post-synaptic potential (EPSP) waveform.  $u(t)$  is the unit step function to ensure that EPSP is zero before the previous spike has occurred.  $E_{syn}$  is the reversal potential,  $i_e$  is the externally applied current,  $i_{app}$  is the externally applied current, and  $netcon$  refers to a binary matrix of network connectivities that define the connections between populations of neurons. Each row in the  $netcon$  matrix represents a presynaptic neuron, and each column represents a postsynaptic neuron. Binary entries of  $netcon$  represents presence of a synaptic connection between neurons. Inhibitory synapses have the following parameters:  $\tau_R = 1ms$ ,  $\tau_D = 10ms$ ,  $E_{syn} = -80mv$ . Excitatory synapses have the following parameters:  $\tau_R = 0.4ms$ ,  $\tau_D = 2ms$ ,  $E_{syn} = 0mv$ . The values for  $g_{syn}$  and  $i_{app}$  are simulation- and connection-dependent, and are listed in Table 2. The network connections are illustrated in Figs 1 and 3.

The default  $g_{syn}$  were chosen such that if I neurons were off, then the inputs would be relayed and combined at the C neuron with a similar firing rate, and if I neurons were on, then E neurons would be completely silenced.

## Simulation-specific model configurations

**Lee & Middlebrooks simulations.** In this spatial tuning simulation, 80 ms of white gaussian noise was placed between  $-80^\circ$  to  $80^\circ$  azimuth, in  $10^\circ$  increments. The spatialized stimuli were then processed and encoded with the subcortical model. The midbrain model in this simulation consisted of 19 spatial channels from  $-90^\circ$  to  $90^\circ$  azimuth, in  $10^\circ$  increments, and 64 frequency channels. To reduce the computational demand of simulating the AIM network, a new set of spike trains, generated using a Poisson model based on the overall firing rate across

**Table 1. Default parameters of cellular dynamics.**

| Parameter              | Value |
|------------------------|-------|
| $C$ (nF)               | 1     |
| $g_{leak}$ ( $\mu S$ ) | 0.1   |
| $E_{leak}$ (mV)        | -70   |
| $V_{thresh}$ (mV)      | -55   |
| $V_{spike}$ (mV)       | 50    |
| $V_{reset}$ (mV)       | -75   |
| $i_{app}$ ( $\mu A$ )  | 0     |
| Noise                  | 0     |

<https://doi.org/10.1371/journal.pcbi.1009356.t001>



**Table 2. Simulation-specific parameters.**  $g_{syn}$  have units of  $\mu S$  and  $i_{app}$  have units of  $nA$ . Parameters of local convergence  $g_{IE}$ ,  $g_{EC}$ ,  $\sigma_{IE}$ , and  $\sigma_{EC}$  are also shown (Fig 3).

| Connection or Neuron        | Param     | Lee | Fritz | Atiani a | Atiani b | Spatial Function | Freq Function |
|-----------------------------|-----------|-----|-------|----------|----------|------------------|---------------|
| $IC \rightarrow E$          | $g_{SYN}$ | 2.5 | 4     | 4        | 3        | 2                | 4             |
| $E \rightarrow C$           | $g_{SYN}$ | 2   | 1.25  | 1.25     | 3        | 2                | 1.5           |
| $E \rightarrow I$           | $g_{SYN}$ | 2.5 | 3     | 3        | 3        | 2                | 3             |
| $E \rightarrow E$ (Attend)  | $g_{SYN}$ | -   | 5     | -        | -        | -                | -             |
| $E \rightarrow E$ (Passive) | $g_{SYN}$ | -   | 3     | -        | -        | -                | -             |
| $I2 \rightarrow I$          | $g_{SYN}$ | 4   | 3     | 3        | 3        | 2.25             | 3             |
| $I2 \rightarrow E$          | $g_{SYN}$ | 2.8 | 4     | 4        | 3        | 1.25             | 3             |
| $I \rightarrow E$           | $g_{SYN}$ | 3   | 4     | 4        | 4        | 3                | 3.5           |
| $I$                         | $i_{app}$ | 3   | -     | -        | -        | 4                | -             |
| $E$                         | $i_{app}$ | 1   | 3     | 3        | 3        | 0                | 3             |
| $I2$                        | $i_{app}$ | 8   | 8     | 8        | 8        | 3.5              | 8             |
| $g_{EC}$                    | -         | -   | 1     | 1        | 1        | -                | 1             |
| $g_{EC}$                    | -         | -   | 1     | 1        | 1        | -                | 1             |
| $\sigma_{IE}$ (kHz)         | -         | -   | 0.65  | 1.7      | 2        | -                | 0.01          |
| $\sigma_{EC}$ (kHz)         | -         | -   | 0.32  | 1.15     | 0.35     | -                | 0.05          |

<https://doi.org/10.1371/journal.pcbi.1009356.t002>

all frequency channels, were computed for each spatial channel. This operation essentially collapses the neural response over the frequency dimension. Therefore, the AIM network for this simulation consisted of 19 spatial channels and one single frequency channel, where each spatial channel processed the set of spike trains that represent the average activity across all frequencies. Network connectivities between spatial channels are as shown in Fig 2. Spatial tuning curves were then calculated based on the response of the C neuron of the AIM network.

The effects of neuromodulators were simulated by applying a gain on the network connections. During the behavior state, off-target E-C connectivities were applied a gain of 0 to simulate the effects of muscarinic receptors, and off-target IC-E connectivities were applied a gain of 2.5 to simulate the effects of the nicotinic receptors. These gains were chosen to replicate the effects observed experimentally.

**Atiani et al. and Fritz et al. simulations.** Pure tones were presented dichotically to the subcortical model, which consisted of a single spatial channel, corresponding to  $0^\circ$  azimuth, and 64 frequency channels. Spike trains were passed directly to the AIM network, which also consisted of a single spatial channel and 64 frequency channels. Network connectivities between spatial channels are as shown in Fig 3. Approximations to spectral temporal receptive fields were calculated based on the response of the best-frequency cortical neuron to each of the pure tone stimulus.

**Calculation of frequency- and azimuth- dependent peristimulus time histograms (PSTHs).** To Approximate STRFs of cortical neurons, we show the responses of the model cortical neuron as functions of time and either frequency or space. In the spatial case, white Gaussian noise were used as stimuli. In the spectral case, pure tones were used as stimuli. Model neuron response for each frequency or azimuth were shown as its firing rate, which was calculated using a 5ms moving window.

**Functional example—spatial listening.** In this example, we demonstrate how the AIM network can be used to isolate a specific talker of interest within a speech mixture. 20 pairs of speech tokens, one male and one female, were randomly chosen from the CRM corpus. The male token was placed at  $0^\circ$  and the female token was placed at  $90^\circ$  azimuth. For simultaneous

presentation, speech tokens were summed prior to being processed by the subcortical model. The subcortical model used 5 spatial channels, tuned from  $-90^\circ$  to  $90^\circ$  azimuth in  $45^\circ$  increments, and 64 frequency channels, and spatially segregates the speech tokens. Its output is relayed directly to the AIM network, which also has 5 spatial channels and 64 frequency channels. In this simulation, network connectivities across spatial channels and parameters are as shown in Fig 2, and each frequency channel operated independently of each other.

To demonstrate the effect of spatial separation on model performance, the male speaker was placed at  $0^\circ$  azimuth while the female speaker was placed at locations between  $15^\circ$  and  $90^\circ$  azimuths in  $15^\circ$  increments (Fig 6E). The subcortical model for this simulation used 7 spatial channels, corresponding to locations from  $0^\circ$  to  $90^\circ$  azimuths in  $15^\circ$  increments, and 64 frequency channels. The output is then processed with the AIM network, which has 7 spatial channels corresponding to the same locations in space. The AIM network was set to 1) attend to the male target at  $0^\circ$  or 2) attend to the female target at various locations or 3) to be in the monitor mode. The network performance was measured by “similarity” between the network outputs and the attended speakers. In the monitoring mode, male speaker was used as the reference talker. Similarity was quantified by calculating the two-dimensional correlation coefficient between the network output of the specific simulation and the network output of the reference speakers. More specifically, we first calculated the firing rates for each frequency channel, then calculated the two-dimensional correlation coefficient of the firing rates.

**Functional example—monaural listening.** In this example, we demonstrate that the AIM network can also operate in the spectral domain to aid in sound segregation during monaural listening. The same 20 pairs of speech tokens as above were summed and presented dichotically to the subcortical model. Here, both the subcortical model and the AIM network has one spatial channel ( $0^\circ$  azimuth) and 64 frequency channels. In this simulation, the network connectivities are as shown in Fig 3. The pitch of each speech token was estimated using MATLAB Audio Toolbox’s `pitch()` function. The I-E connectivity parameters were chosen based on the two speaker’s  $f_0$ , such that when attention is focused on one speaker’s  $f_0$ , the other speaker’s  $f_0$  would be inhibited.

## Supporting information

**S1 Fig. Two possible modes to achieve top-down lateral inhibition, resulting in sharpening of the spatial tuning of the attended channel.** A) Direct inhibition from I2 to I neuron, and B) Feed forward inhibition from I2  $\rightarrow$  E  $\rightarrow$  I neuron. Voltage traces of each model neuron under the passive or attending condition is shown on first two columns of the grid. The final column shows spatial tuning of the attended channel in the passive vs attending conditions. (TIF)

**S2 Fig. AIM mechanism can weaken inhibitory regions of frequency-dependent PSTHs.** A) Within-channel inhibition (S neuron) is added to the AIM network for this simulation to induce suppression of activity relative to spontaneous firing of E neurons, thereby creating an inhibitory region in the frequency-dependent PSTH. Note that S neurons are distinct from I neurons, which inhibit other frequency channels. B) Frequency-dependent PSTHs. Blue regions in PSTH indicate lower firing rate relative to the spontaneous firing rate. When attention is turned on and I2 neuron is turned off, the E neuron is released from inhibition, resulting in the weakening of the inhibitory region. Simulation parameters are listed in the table below. (TIF)

**S1 Table. Model parameters for attention-induced weakening of inhibitory regions in frequency-dependent PSTHs as shown in S2 Fig.**  
(XLSX)

## Acknowledgments

The authors thank Larry Abbott and John Middlebrooks for comments on the manuscript.

## Author Contributions

**Conceptualization:** Kamal Sen.

**Formal analysis:** Kenny F. Chou.

**Funding acquisition:** Kamal Sen.

**Methodology:** Kenny F. Chou.

**Software:** Kenny F. Chou.

**Supervision:** Kamal Sen.

**Validation:** Kenny F. Chou.

**Visualization:** Kenny F. Chou.

**Writing – original draft:** Kenny F. Chou, Kamal Sen.

**Writing – review & editing:** Kenny F. Chou, Kamal Sen.

## References

1. McDermott JH. The cocktail party problem. *Curr Biol.* 2009; 19: R1024–R1027. <https://doi.org/10.1016/j.cub.2009.09.005> PMID: 19948136
2. Vogels TP, Abbott LF. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat Neurosci.* 2009; 12: 483–491. <https://doi.org/10.1038/nn.2276> PMID: 19305402
3. Yang GR, Murray JD, Wang XJ. A dendritic disinhibitory circuit mechanism for pathway-specific gating. *Nat Commun.* 2016;7. <https://doi.org/10.1038/ncomms12815> PMID: 27649374
4. Wang XJ, Yang GR. A disinhibitory circuit motif and flexible information routing in the brain. *Curr Opin Neurobiol.* 2018; 49: 75–83. <https://doi.org/10.1016/j.conb.2018.01.002> PMID: 29414069
5. Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang W-C, Jenvay S, et al. Long-range and local circuits for top-down modulation of visual cortex processing. *Science (80-).* 2014; 345: 660–665. <https://doi.org/10.1126/science.1254126> PMID: 25104383
6. Letzkus JJ, Wolff SBE, Meyer EMM, Tovote P, Courtin J, Herry C, et al. A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature.* 2011; 480: 331–335. <https://doi.org/10.1038/nature10674> PMID: 22158104
7. Pi HJ, Hangya B, Kvitsiani D, Sanders JL, Huang ZJ, Kepecs A. Cortical interneurons that specialize in disinhibitory control. *Nature.* 2013; 503: 521–524. <https://doi.org/10.1038/nature12676> PMID: 24097352
8. Kuchibhotla K V., Gill J V., Lindsay GW, Papadoyannis ES, Field RE, Sten TAH, et al. Parallel processing by cortical inhibition enables context-dependent behavior. *Nat Neurosci.* 2017; 20: 62–71. <https://doi.org/10.1038/nn.4436> PMID: 27798631
9. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience.* 1995. <https://doi.org/10.1146/annurev.ne.18.030195.001205> PMID: 7605061
10. Reynolds JH, Heeger DJ. The Normalization Model of Attention. *Neuron.* 2009. pp. 168–185. <https://doi.org/10.1016/j.neuron.2009.01.002> PMID: 19186161
11. Hubel DH, Henson CO, Rupert A, Galambos R. “Attention” Units in the Auditory Cortex. *Science (80-).* 1959; 129: 1279–1280. <https://doi.org/10.1126/science.129.3358.1279> PMID: 13658956
12. Fritz JB, Elhilali M, David S V, Shamma SA. Auditory attention—focusing the searchlight on sound. *Curr Opin Neurobiol.* 2007; 17: 437–455. <https://doi.org/10.1016/j.conb.2007.07.011> PMID: 17714933

13. Otazu GH, Tai LH, Yang Y, Zador AM. Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci*. 2009. <https://doi.org/10.1038/nn.2306> PMID: 19363491
14. Shinn-Cunningham BG. Object-based auditory and visual attention. *Trends Cogn Sci*. 2008. <https://doi.org/10.1016/j.tics.2008.02.003> PMID: 18396091
15. Buffalo EA, Fries P, Landman R, Liang H, Desimone R. A backward progression of attentional effects in the ventral stream. *Proc Natl Acad Sci U S A*. 2010. <https://doi.org/10.1073/pnas.0907658106> PMID: 20007766
16. Lee C-CC, Middlebrooks JC. Auditory cortex spatial sensitivity sharpens during task performance. *Nat Neurosci*. 2011; 14: 108–114. <https://doi.org/10.1038/nn.2713> PMID: 21151120
17. Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*. 2003; 6: 1216–1223. <https://doi.org/10.1038/nn1141> PMID: 14583754
18. Maddox RK, Billimoria CP, Perrone BP, Shinn-Cunningham BG, Sen K. Competing sound sources reveal spatial effects in cortical processing. *PLoS Biol*. 2012;10. <https://doi.org/10.1371/journal.pbio.1001319> PMID: 22563301
19. Middlebrooks JC, Bremen P. Spatial Stream Segregation by Auditory Cortical Neurons. *J Neurosci*. 2013; 33: 10986–11001. <https://doi.org/10.1523/JNEUROSCI.1065-13.2013> PMID: 23825404
20. Bee MA, Micheyl C. The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol*. 2008; 122: 235–251. <https://doi.org/10.1037/0735-7036.122.3.235> PMID: 18729652
21. Bronkhorst AW. The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, Perception, and Psychophysics*. 2015. pp. 1465–1487. <https://doi.org/10.3758/s13414-015-0882-9> PMID: 25828463
22. Kaya EM, Elhilali M. Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society; 2017. <https://doi.org/10.1098/rstb.2016.0101> PMID: 28044012
23. Kayser C, Petkov CI, Lippert M, Logothetis NK. Mechanisms for allocating auditory attention: An auditory saliency map. *Curr Biol*. 2005; 15: 1943–1947. <https://doi.org/10.1016/j.cub.2005.09.040> PMID: 16271872
24. Kaya EM, Elhilali M. Investigating bottom-up auditory attention. *Front Hum Neurosci*. 2014; 8: 327. <https://doi.org/10.3389/fnhum.2014.00327> PMID: 24904367
25. Mesgarani N, Fritz J, Shamma S. A computational model of rapid task-related plasticity of auditory cortical receptive fields. *J Comput Neurosci*. 2010; 28: 19–27. <https://doi.org/10.1007/s10827-009-0181-3> PMID: 19711179
26. Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. *Neuron*. 2009; 61: 317–329. <https://doi.org/10.1016/j.neuron.2008.12.005> PMID: 19186172
27. Carlin MA, Elhilali M. Modeling attention-driven plasticity in auditory cortical receptive fields. *Front Comput Neurosci*. 2015; 9: 106. <https://doi.org/10.3389/fncom.2015.00106> PMID: 26347643
28. David S V., Fritz JB, Shamma SA. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci U S A*. 2012; 109: 2144–2149. <https://doi.org/10.1073/pnas.1117717109> PMID: 22308415
29. Dong J, Colburn HS, Sen K. Cortical Transformation of Spatial Processing for Solving the Cocktail Party Problem: A Computational Model. *eNeuro*. 2016; 3: 1–11. <https://doi.org/10.1523/ENEURO.0086-15.2015> PMID: 26866056
30. Chou KF, Dong J, Colburn HS, Sen K. A Physiologically Inspired Model for Solving the Cocktail Party Problem. *J Assoc Res Otolaryngol*. 2019; 20: 579–593. <https://doi.org/10.1007/s10162-019-00732-4> PMID: 31392449
31. Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M. Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nat Neurosci*. 2013; 16: 1068–1076. <https://doi.org/10.1038/nn.3446> PMID: 23817549
32. Gil Z, Connors BW, Amitai Y. Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron*. 1997. [https://doi.org/10.1016/s0896-6273\(00\)80380-3](https://doi.org/10.1016/s0896-6273(00)80380-3) PMID: 9331357
33. Hasselmo ME. Neuromodulation and cortical function: modeling the physiological basis of behavior. *Behav Brain Res*. 1995. [https://doi.org/10.1016/0166-4328\(94\)00113-t](https://doi.org/10.1016/0166-4328(94)00113-t) PMID: 7748496
34. Hsieh CY, Cruikshank SJ, Metherate R. Differential modulation of auditory thalamocortical and intracortical synaptic transmission by cholinergic agonist. *Brain Res*. 2000. [https://doi.org/10.1016/s0006-8993\(00\)02766-9](https://doi.org/10.1016/s0006-8993(00)02766-9) PMID: 11032989

35. Atiani S, Elhilali M, David S V., Fritz JB, Shamma SA. Task Difficulty and Performance Induce Diverse Adaptive Patterns in Gain and Shape of Primary Auditory Cortical Receptive Fields. *Neuron*. 2009; 61: 467–480. <https://doi.org/10.1016/j.neuron.2008.12.027> PMID: 19217382
36. Litovsky RY. Spatial Release from Masking. *Acoust Today*. 2012; 8: 18. <https://doi.org/10.1121/1.4729575>
37. Zhang S, Xu M, Kamigaki T, Do JPH, Chang WC, Jenvey S, et al. Long-range and local circuits for top-down modulation of visual cortex processing. *Science (80-)*. 2014; 345: 660–665. <https://doi.org/10.1126/science.1254126> PMID: 25104383
38. Helfrich RF, Fiebelkorn IC, Szczepanski SM, Lin JJ, Parvizi J, Knight RT, et al. Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron*. 2018. <https://doi.org/10.1016/j.neuron.2018.07.032> PMID: 30138591
39. Guo W, Clause AR, Barth-Maroon A, Polley DB. A Corticothalamic Circuit for Dynamic Switching between Feature Detection and Discrimination. *Neuron*. 2017. <https://doi.org/10.1016/j.neuron.2017.05.019> PMID: 28625486
40. Huang N, Elhilali M. Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. *Elife*. 2020. <https://doi.org/10.7554/eLife.52984> PMID: 32196457
41. Fiebelkorn IC, Kastner S. A Rhythmic Theory of Attention. *Trends in Cognitive Sciences*. 2019. <https://doi.org/10.1016/j.tics.2018.11.009> PMID: 30591373
42. Lee SH, Dan Y. Neuromodulation of Brain States. *Neuron*. 2012; 76: 209–222. <https://doi.org/10.1016/j.neuron.2012.09.012> PMID: 23040816
43. Lesenfants D, Francart T. The interplay of top-down focal attention and the cortical tracking of speech. *Sci Rep*. 2020; 10: 1–10. <https://doi.org/10.1038/s41598-019-56847-4> PMID: 31913322
44. Wikman P, Sahari E, Salmela V, Leminen A, Leminen M, Laine M, et al. Breaking down the cocktail party: Attentional modulation of cerebral audiovisual speech processing. *Neuroimage*. 2021; 224: 117365. <https://doi.org/10.1016/j.neuroimage.2020.117365> PMID: 32941985
45. Gritton HJ, Nocon JC, James NM, Lowet E, Abdulkarim M, Sen K, et al. Oscillatory activity in alpha/beta frequencies coordinates auditory and prefrontal cortices during extinction learning. *bioRxiv*. 2020; 2020.10.30.362962. <https://doi.org/10.1101/2020.10.30.362962>
46. Corbetta M, Patel G, Shulman GL. The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*. NIH Public Access; 2008. pp. 306–324. <https://doi.org/10.1016/j.neuron.2008.04.017> PMID: 18466742
47. Petersen SE, Posner MI. The Attention System of the Human Brain: 20 Years After. *Annu Rev Neurosci*. 2012; 35: 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525> PMID: 22524787
48. Tremblay R, Lee S, Rudy B. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*. 2016; 91: 260–292. <https://doi.org/10.1016/j.neuron.2016.06.033> PMID: 27477017
49. Karnani MM, Jackson J, Ayzenshtat I, Sichani XH, Manoocheri K, Kim S, et al. Opening holes in the blanket of inhibition: Localized lateral disinhibition by vip interneurons. *J Neurosci*. 2016; 36: 3471–3480. <https://doi.org/10.1523/JNEUROSCI.3646-15.2016> PMID: 27013676
50. Pfeffer CK. Inhibitory Neurons: Vip Cells Hit the Brake on Inhibition. *Curr Biol*. 2014; 24: R18–R20. <https://doi.org/10.1016/j.cub.2013.11.001> PMID: 24405670
51. Muñoz W, Tremblay R, Levenstein D, Rudy B. Layer-specific modulation of neocortical dendritic inhibition during active wakefulness. *Science (80-)*. 2017. <https://doi.org/10.1126/science.aag2599> PMID: 28254942
52. Silberberg G, Markram H. Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. *Neuron*. 2007. <https://doi.org/10.1016/j.neuron.2007.02.012> PMID: 17329212
53. Naka A, Veit J, Shababo B, Chance RK, Rizzo D, Stafford D, et al. Complementary networks of cortical somatostatin interneurons enforce layer specific control. *Elife*. 2019. <https://doi.org/10.7554/eLife.43696> PMID: 30883329
54. van der Heijden K, Rauschecker JP, Formisano E, Valente G, de Gelder B. Active Sound Localization Sharpens Spatial Tuning in Human Primary Auditory Cortex. *J Neurosci*. 2018; 38: 8574–8587. <https://doi.org/10.1523/JNEUROSCI.0587-18.2018> PMID: 30126968
55. Fritz JB, Elhilali M, David S V., Shamma SA. Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear Res*. 2007; 229: 186–203. <https://doi.org/10.1016/j.heares.2007.01.009> PMID: 17329048
56. Panniello M, King AJ, Dahmen JC, Walker KMM. Local and global spatial organization of interaural level difference and frequency preferences in auditory cortex. *Cereb Cortex*. 2018; 28: 350–369. <https://doi.org/10.1093/cercor/bhx295> PMID: 29136122

57. Knudsen E, Konishi M. A neural map of auditory space in the owl. *Science* (80-). 1978; 200: 795–797. <https://doi.org/10.1126/science.644324> PMID: 644324
58. Yin TCT, Chan JCK. Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol.* 1990; 64: 465–488. <https://doi.org/10.1152/jn.1990.64.2.465> PMID: 2213127
59. Köppl C, Carr CE. Maps of interaural time difference in the chicken's brainstem nucleus laminaris. *Biol Cybern.* 2008; 98: 541–559. <https://doi.org/10.1007/s00422-008-0220-6> PMID: 18491165
60. Zhou Y, Wang X. Spatially extended forward suppression in primate auditory cortex. *Eur J Neurosci.* 2014; 39: 919–933. <https://doi.org/10.1111/ejn.12460> PMID: 24372934
61. Yao JD, Bremen P, Middlebrooks JC. Emergence of Spatial Stream Segregation in the Ascending Auditory Pathway. *J Neurosci.* 2015; 35: 16199–16212. <https://doi.org/10.1523/JNEUROSCI.3116-15.2015> PMID: 26658870
62. Higgins NC, Storace DA, Escabi MA, Read HL. Specialization of Binaural Responses in Ventral Auditory Cortices. *J Neurosci.* 2010; 30: 14522–14532. <https://doi.org/10.1523/JNEUROSCI.2561-10.2010> PMID: 20980610
63. Kyweriga M, Stewart W, Cahill C, Wehr M. Synaptic mechanisms underlying interaural level difference selectivity in rat auditory cortex. *J Neurophysiol.* 2014; 112: 2561–2571. <https://doi.org/10.1152/jn.00389.2014> PMID: 25185807
64. Middlebrooks J, Knudsen E. A neural code for auditory space in the cat's superior colliculus. *J Neurosci.* 1984; 4: 2621–2634. <https://doi.org/10.1523/JNEUROSCI.04-10-02621.1984> PMID: 6491727
65. King AJ, Hutchings ME. Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: A map of auditory space. *J Neurophysiol.* 1987; 57: 596–624. <https://doi.org/10.1152/jn.1987.57.2.596> PMID: 3559694
66. Chou XL, Fang Q, Yan L, Zhong W, Peng B, Li H, et al. Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression. *Elife.* 2020; 9: 1–21. <https://doi.org/10.7554/eLife.54157> PMID: 32142411
67. Narayan R, Ergün A, Sen K. Delayed inhibition in cortical receptive fields and the discrimination of complex stimuli. *J Neurophysiol.* 2005; 94: 2970–2975. <https://doi.org/10.1152/jn.00144.2005> PMID: 15917327
68. Feng L, Wang X. Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proc Natl Acad Sci U S A.* 2017; 114: E840–E848. <https://doi.org/10.1073/pnas.1607519114> PMID: 28096341
69. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature.* 2012; 485: 233–236. <https://doi.org/10.1038/nature11020> PMID: 22522927
70. Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci.* 2012; 109: 11854–11859. <https://doi.org/10.1073/pnas.1205381109> PMID: 22753470
71. Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 2011; 34: 114–123. <https://doi.org/10.1016/j.tins.2010.11.002> PMID: 21196054
72. Krishnan L, Elhilali M, Shamma S. Segregating Complex Sound Sources through Temporal Coherence. Lewicki M, editor. *PLoS Comput Biol.* 2014; 10: e1003985. <https://doi.org/10.1371/journal.pcbi.1003985> PMID: 25521593
73. Chambers JD, Elgueda D, Fritz JB, Shamma SA, Burkitt AN, Grayden DB. Computational neural modeling of auditory cortical receptive fields. *Front Comput Neurosci.* 2019; 13: 1–13. <https://doi.org/10.3389/fncom.2019.00001> PMID: 30809141
74. Soto G, Kopell N, Sen K. Network architecture, receptive fields, and neuromodulation: Computational and functional implications of cholinergic modulation in primary auditory cortex. *J Neurophysiol.* 2006; 96: 2972–2983. <https://doi.org/10.1152/jn.00459.2006> PMID: 16899641
75. Ardid S, Wang XJ, Compte A. An integrated microcircuit model of attentional processing in the neocortex. *J Neurosci.* 2007. <https://doi.org/10.1523/JNEUROSCI.1145-07.2007> PMID: 17687026
76. Bolia RS, Nelson WT, Ericson MA, Simpson BD. A speech corpus for multitalker communications research. *J Acoust Soc Am.* 2000; 107: 1065–1066. <https://doi.org/10.1121/1.428288> PMID: 10687719
77. Algazi VR, Duda RO, Thompson DM, Avendano C. The CIPIC HRTF database. *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat No01TH8575).* IEEE; 2001. pp. 99–102. <https://doi.org/10.1109/ASPAA.2001.969552>
78. Burkhard MD, Sachs RM. Anthropometric manikin for acoustic research. *J Acoust Soc Am.* 1975; 58: 214–222. <https://doi.org/10.1121/1.380648> PMID: 1150969

79. Slaney M. Auditory toolbox: A Matlab Toolbox for Auditory Modeling Work. Interval Res Corp Tech Rep. 1998; 10: 1998. Available: <https://engineering.purdue.edu/~malcolm/interval/1998-010/AuditoryToolboxTechReport.pdf>
80. Fischer BJ, Anderson CH, Peña JL. Multiplicative auditory spatial receptive fields created by a hierarchy of population codes. PLoS One. 2009; 4: 24–26. <https://doi.org/10.1371/journal.pone.0008015> PMID: 19956693
81. Sherfey JS, Soplata AE, Ardid S, Roberts EA, Stanley DA, Pittman-Polletta BR, et al. DynaSim: A MATLAB Toolbox for Neural Modeling and Simulation. Front Neuroinform. 2018; 12: 10. <https://doi.org/10.3389/fninf.2018.00010> PMID: 29599715
82. Dayan P, Abbott LF. Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT Press; 2001.