Differential contributions of synaptic and intrinsic inhibitory currents to speech segmentation via flexible phase-locking in neural oscillators

Benjamin R. Pittman-Polletta[1*], Yangyang Wang[2], David A. Stanley[1], Charles E. Schroeder[3], Miles A. Whittington[4], Nancy J. Kopell[1],

1 Department of Mathematics & Statistics, Boston University, Boston, MA **2** Department of Mathematics, University of Iowa, Iowa City, IA **3** Translational Neuroscience Division, Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA **4** Hull York Medical School, York, UK

* benpolletta@gmail.com

Abstract

Current hypotheses suggest that speech segmentation – the initial division and grouping of the speech stream into candidate phrases, syllables, and phonemes for further linguistic processing – is executed by a hierarchy of oscillators in auditory cortex. Theta (∼3-12 Hz) rhythms play a key role by phase-locking to recurring acoustic features marking syllable boundaries. Reliable synchronization to quasi-rhythmic inputs, whose variable frequency can dip below cortical theta frequencies (down to ∼1 Hz), requires "flexible" theta oscillators whose underlying neuronal mechanisms remain unknown. Using biophysical computational models, we found that the flexibility of phase-locking in neural oscillators depended on the types of hyperpolarizing currents that paced them. Simulated cortical theta oscillators flexibly phase-locked to slow inputs when these inputs caused both (i) spiking and (ii) the subsequent buildup of outward current sufficient to delay further spiking until the next input. The greatest flexibility in phase-locking arose from a synergistic interaction between intrinsic currents that was not replicated by synaptic currents at similar timescales. Flexibility in phase-locking enabled improved entrainment to speech input, optimal at mid-vocalic channels, which

in turn supported syllabic-timescale segmentation through identification of vocalic nuclei. Our results suggest that synaptic and intrinsic inhibition contribute to frequency-restricted and -flexible phase-locking in neural oscillators, respectively. Their differential deployment may enable neural oscillators to play diverse roles, from reliable internal clocking to adaptive segmentation of quasi-regular sensory inputs like speech.

Author summary

Oscillatory activity in auditory cortex is believed to play an important role in auditory and speech processing. One suggested function of these rhythms is to divide the speech stream into candidate phonemes, syllables, words, and phrases, to be matched with learned linguistic templates. This requires brain rhythms to flexibly synchronize with regular acoustic features of the speech stream. How neuronal circuits implement this task remains unknown. In this study, we explored the contribution of inhibitory currents to flexible phase-locking in neuronal theta oscillators, believed to perform initial syllabic segmentation. We found that a combination of specific intrinsic inhibitory currents at multiple timescales, present in a large class of cortical neurons, enabled exceptionally flexible phase-locking, which could be used to precisely segment speech by identifying vowels at mid-syllable. This suggests that the cells exhibiting these currents are a key component in the brain's auditory and speech processing architecture.

1 Introduction

Conventional models of speech processing [1–3] suggest that decoding proceeds by matching chunks of speech of different durations with stored linguistic memory patterns or templates. Recent oscillation-based models have postulated that this template-matching is facilitated by a preliminary segmentation step [4–8], which determines candidate speech segments for template matching, in the process tracking speech speed and allowing the adjustment (within limits) of sampling and segmentation rates [9, 10]. Segmentation plays a key role in explaining a range of counterintuitive psychophysical data that challenge conventional models of speech perception [8, 11–13], and conceptual hypotheses [6, 7, 14–18] suggest cortical rhythms entrain to regular

acoustic features of the speech stream [19–22] to effect this preliminary grouping of auditory input.

Speech is a multiscale phenomenon, but both the amplitude modulation of continuous speech and the motor physiology of the speech apparatus are dominated by syllabic timescales – i.e., δ/θ frequencies ($\sim$1-9 Hz) [23–27]. This syllabic timescale information is critical for speech comprehension [11, 12, 26, 28–31], as is speech-brain entrainment at δ/θ frequencies [32–38], which may play a causal role in speech perception [39–42]. Cortical θ rhythms – especially prominent in the spontaneous activity of primate auditory cortex [43] – seem to perform an essential function in syllable segmentation [11–13, 37], and seminal phenomenological [11] and computational [44–47] models have proposed a framework in which putative syllables segmented by θ oscillators drive speech sampling and encoding by γ ($\sim$30-60 Hz) oscillatory circuits. The fact that oscillator-based syllable boundary detection performs better than classical algorithms [45, 46] argues for the role of endogenous rhythmicity – as opposed to merely event-related responses to rhythmic inputs – in speech segmentation and perception.

However, there are issues with existing models. *In vitro* results show that the dynamics of cortical, as opposed to hippocampal [48], θ oscillators depend on intrinsic currents at least as much as (and arguably more than) synaptic currents [49, 50]. Yet existing models of oscillatory syllable segmentation assume θ rhythms are paced by synaptic inhibition [45, 47], and employ methodologies – integrate-and-fire neurons [45] or one-dimensional oscillators [47] – incapable of capturing the dynamics of intrinsic currents. This is important because the variability of syllable lengths between syllables, speakers, and languages, as well as across linguistic contexts, demands "flexibility" – the ability to phase-lock, cycle-by-cycle, to quasi-rhythmic inputs with a broad range of instantaneous frequencies [6, 12], including those below its intrinsic frequency – of any cortical θ oscillator tasked with syllabic segmentation. In contrast to this functional constraint, (synaptic) inhibition-based rhythms have been shown to exhibit *inflexibility* in phase-locking, especially to input frequencies lower than their intrinsic frequency [51, 52]. Furthermore, the pattern of spiking exhibited by a flexible θ rhythm – which we show depends markedly on the intrinsic currents it exhibits – has important implications for downstream speech processing, being hypothesized to determine how

and at what speed β- ($\sim$15-30 Hz) and γ-rhythmic cortical circuits sample and predict acoustic information [47, 53]. And while much is known about phase-locking in neural oscillators [54–58], the existing literature sheds little light on these issues: few studies have examined the physiologically relevant "strong forcing regime", in which input pulses are strong enough to elicit spiking [59]; little work has explored how oscillator parameters influence phase-locking to inputs much slower or faster than an oscillator's intrinsic frequency [60]; and few published studies explore oscillators exhibiting intrinsic outward currents on multiple timescales [61].

In addition, the lack of reliable acoustic markers of syllable boundaries, and the higher information density of the consonantal clusters that mark *linguistic* syllable boundaries relative to the high energy and long-duration vowels at their center, has led to the suggestion that reliable speech-brain entrainment may reverse the syllabic convention, relying on the high energy vocalic nuclei at the center of each syllable to mark segmental boundaries [6, 12, 16, 45, 62], enabling both robust determination of these boundaries and dependable sampling of the consonantal evidence that informs segment identity. These reversed "theta-syllables" are hypothesized to be the candidate cortical segments distinguished and passed downstream for further processing [16] by auditory cortical θ rhythms, but whether θ rhythms differentially entrain to different speech channels (associated with the acoustics of consonants and vowels), and the impact of such differential entrainment on speech segmentation at the syllabic timescale, remain unexamined.

Motivated by these issues, we explored whether and how the biophysical mechanisms giving rise to cortical θ oscillations affect their ability to flexibly phase-lock to inputs containing frequencies slower than their intrinsic frequency. We tested the phase-locking capabilities of biophysical computational models of neural θ oscillators, parameterized to spike intrinsically at 7 Hz, and containing all feasible combinations of: (i) θ-timescale subthreshold oscillations (STOs) resulting from an intrinsic θ-timescale hyperpolarizing current (as observed in θ-rhythmic layer 5 pyramids [50, 63], and whose presence is denoted by "M" in the name of the model); (ii) an intrinsic "super-slow" (δ-timescale) hyperpolarizing current (also observed *in vitro* [50], and present in models with an "S"); and (iii) θ-timescale synaptic inhibition, as previously modeled [45] (present in models with an "I"). We drove these oscillators with synthetic periodic and quasi-periodic

inputs, as well as speech inputs derived from the TIMIT corpus [64]. To determine whether and how these oscillators' spiking activity could contribute to meaningful syllabic-timescale segmentation, we used speech-driven model spiking to derive putative segmental boundaries, and compared these boundaries' temporal and phonemic distribution to syllabic midpoints obtained from phonemic transcriptions.

Models exhibiting the combination of STOs and super-slow rhythms observed *in vitro* (models MS and MIS) showed markedly more flexible phase-locking to synthetic inputs than primarily inhibition-paced models (models I, MI, and IS), and yielded segmental boundaries closer to syllabic midpoints, even when phase-locking to speech was hampered by a higher overall level of inhibition (model MIS). Exploring the activation of these three inhibitory currents immediately prior to spiking revealed that flexible phase-locking was driven by a novel complex interaction between θ-timescale STOs and super-slow K currents. This interaction enabled a buildup of outward (inhibitory) current during input pulses, sufficiently long-lasting to silence spiking during the period between successive inputs, even when this period lasted for many θ cycles, that was absent from oscillators paced by synaptic inhibition. All our models phase-locked most strongly to mid-vocalic channels and produced segmental boundaries predominately during vocalic phonemes, supporting the notion that θ-rhythmic syllable segmentation may make use of θ-syllables rather than conventional, linguistically defined ones.

2 Results

2.1 Modeling cortical θ oscillators

To investigate the dependence of frequency flexibility in phase-locking on the biophysics and dynamics of inhibitory currents, we employed Hodgkin-Huxley type computational models of cortical θ oscillators. In these models, θ rhythmicity was paced by either or both of two mechanisms: synaptic inhibition with a fast rise time and a slow decay time as in the hippocampus [48] and previous models of syllable segmentation [45]; and θ-frequency sub-threshold oscillations (STOs) resulting from the interaction of a pair of intrinsic currents activated at subthreshold membrane potentials – a depolarizing persistent sodium current and a hyperpolarizing and slowly activating m-current [49]

(Fig. 1, 2B). A super-slow potassium current introduced a δ timescale into the dynamics of some models (Fig. 1, 2C) and helped to recreate dynamics observed *in vitro* [50]. Thus, in addition to spiking and leak currents, our models included up to three types of outward – i.e. hyperpolarizing and thus spike suppressing, and here termed inhibitory – currents: an m-current or slow potassium current (I_{m}) with a voltage-dependent time constant of activation of ∼10-45 ms; recurrent synaptic inhibition (I_{inh}) with a decay time of 60 ms; and a super-slow K current ($I_{\mathrm{K_{SS}}}$) with (calcium-dependent) rise and decay times of ∼100 and ∼500 ms, respectively. The presence of these three hyperpolarizing currents was varied over six models – M, I, MI, MS, IS, and MIS – whose names indicate the presence of each current: M for the m-current, I for synaptic inhibition, and S for the super-slow K current (Fig. 1).

We began by qualitatively matching *in vitro* recordings from layer 5 θ-resonant pyramidal cells [50] (Fig. 2D). As their resting membrane potential is raised over a few mV, these RS cells exhibit a characteristic transition from tonic δ-rhythmic spiking to tonic θ-rhythmic spiking through so-called mixed-mode oscillations (MMOs, here doublets of spikes spaced a θ period apart occurring at a δ frequency) [50]. *In vitro* data suggests that this pattern of spiking is independent of recurrent synaptic inhibition, arising instead from intrinsic inhibitory currents. To replicate this behavior, we constructed a Hodgkin-Huxley neuron model paced by both I_{m} and $I_{\mathrm{K_{SS}}}$ (Fig. 1, model MS, & Fig. 2A). While *in vitro*, these layer 5 θ-rhythmic pyramidal cells receive δ-rhythmic EPSPs, this rhythmic excitation is not required in our model, which exhibited MMOs in response to tonic input (Fig. 2D).

We then constructed five additional models based on model MS (Fig. 1). To compare the performance of this model to inhibition-based oscillators, we obtained model IS by replacing I_{m} with feedback synaptic inhibition I_{inh} from a SOM-like interneuron (Fig. 1, model IS, schematic), adjusting the leak current and the conductance of synaptic inhibition to get a frequency-current (FI) curve having a rheobase and inflection point similar to that of model MS (Fig. 1, model IS, FI curve). In the remaining models, only the leak current conductance was changed, to enable 7 Hz tonic spiking at roughly similar values of I_{app}; except for the presence of the three inhibitory currents, all other conductances were identical to those in models MS and IS (see Table 3). Two models without $I_{\mathrm{K_{SS}}}$ (model M and model I) were constructed to explore this current's

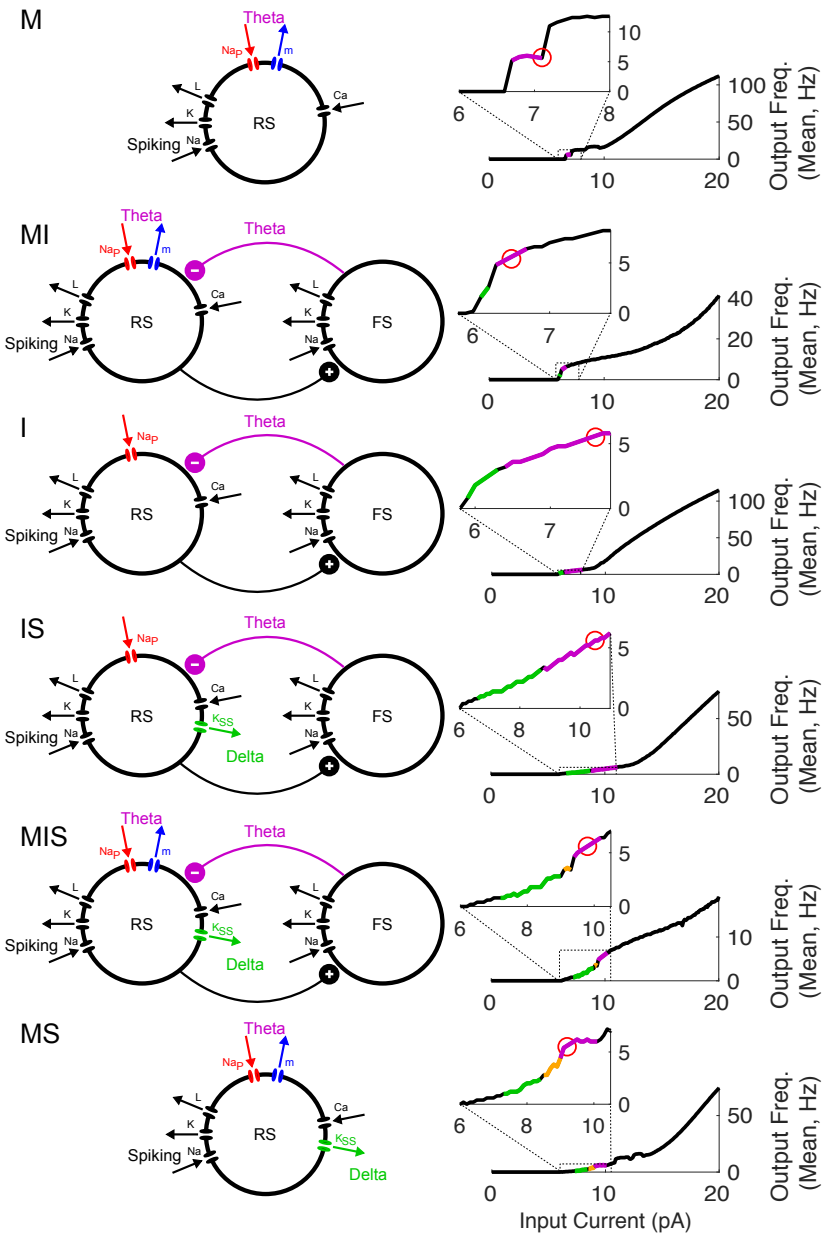

Fig 1. Model θ oscillators. Schematics (left) show the currents present in each model, color-coded according to timescale of inhibition (δ in green, θ in purple). FI curves (right) show the transition of spiking rhythmicity through δ and θ frequencies as I_{app} increases (δ in green, θ in purple, and MMOs in gold); the red circle indicates the point on the FI curve at which I_{app} was fixed, to give a 7 Hz firing rate.

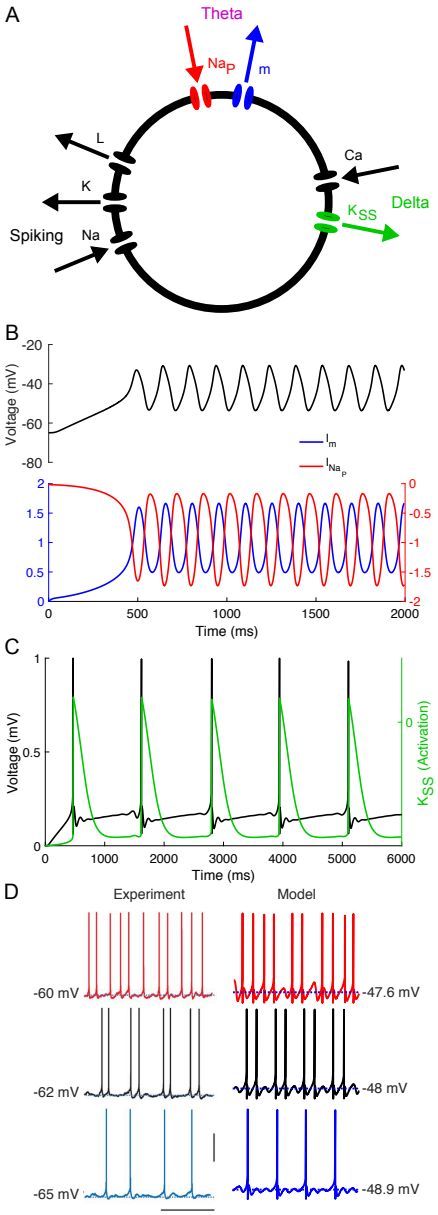

Fig 2. Model MS reproduces *in vitro* data. (A) Diagram of model MS. Arrows indicate directions of currents (i.e., inward or outward). (B) θ timescale STOs arise from interactions between m- and persistent sodium currents in a model without spiking or Ca-dependent currents (only g_{m} and g_{Na_P} nonzero). (C) δ timescale activity-dependent hyperpolarization arises from a super-slow K current. (D) Comparison between *in vitro* (adapted from [50]) and model data (vertical bar 50 μV, horizontal bar 0.5 ms).

contribution to model phase-locking. Two more models were constructed with both I_{m} and I_{inh} to explore the interactions of these currents. (Models with neither I_{m} nor I_{inh} lacked robust 7 Hz spiking.) For all simulations, we chose and fixed I_{app} so that all models exhibited intrinsic rhythmicity at the same frequency, 7 Hz (Fig. 1, inset FI curves, red circles), allowing us to directly compare the frequency range of phase-locking between models.

2.2 Phase-locking to strong forcing by simulated inputs

We tested the entrainment of these model oscillators using simulated inputs strong enough to cause spiking with each input "pulse".

2.2.1 Rhythmic inputs

To begin mapping the frequency range of phase-locking in our models, we measured model phase-locking in response to regular rhythmic inputs, modeled as smoothed square-wave current injections to the RS cells of all three models. The frequencies of these inputs ranged from 0.25 to 23 Hz, and their duty cycles were held constant at 1/4 of the input period (see Methods), to mimic the bursts of excitation produced by deep intrinsic bursting (IB) cells projecting to deep regular spiking (RS) cells [50]. For inputs at all frequencies, the total (integrated) input over 30 s was normalized, and multiplied by a gain varied from 0 to 4. Entrainment was measured as the phase-locking value (PLV) of RS cell spiking to the input rhythm phase (see Section 4.3).

The results of these simulations are shown in Fig. 3, with models ordered by increasing frequency flexibility of phase-locking, as measured by the lower frequency limit of appreciable phase-locking. The most flexible model (MS) was able to phase lock to input frequencies as low as 1.5 Hz even when input strength was relatively low, while the least flexible model (M) was unable to phase-lock to input frequencies below 7 Hz. For high enough input strength, all models were able to phase-lock adequately to inputs faster than 7 Hz, up to and including the fastest frequency we tested (23 Hz). However, much of this phase-locking occurred with less than one spike per input cycle (see white contours, Fig. 3). Notably, models MI and MIS exhibited one-to-one phase-locking to periodic inputs at input strengths twice as high as other models. Simulations showed

that this was due to a higher overall level of inhibition, as the range of input strengths over which one-to-one phase-locking was observed increased with the conductances of both the m- and synaptic currents (Fig. S1).

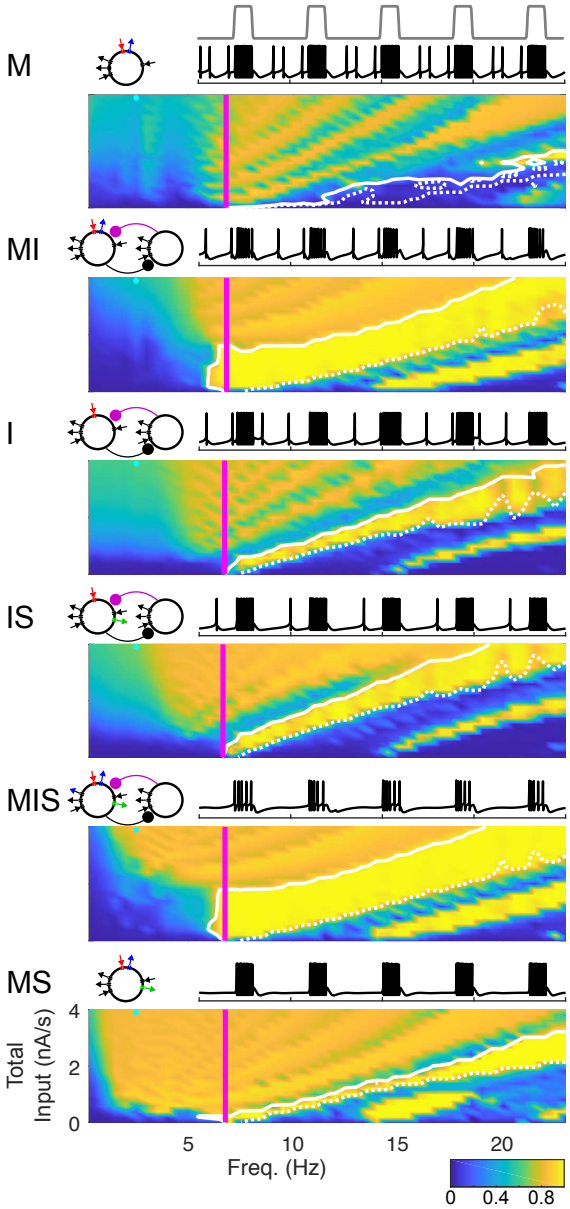

Fig 3. Phase-Locking as a Function of Periodic Input Frequency & Strength. False-color images show the (spike-rate adjusted) phase-locking value (PLV, see Section 4.3) of spiking to input waveform. Vertical magenta lines indicating intrinsic spiking frequency. Solid white contour indicates boundary of phase-locking with one spike per cycle; dotted white contour indicates boundary of phase-locking with 0.9 spikes per cycle. Bands in false-color images of PLV are related to the number of spikes generated per input cycle: the highest PLV occurs when an oscillator produces one spike per input cycle, and PLV decreases slightly (from band to band) as both the strength of the input and the number of spikes per input cycle increases. Schematics of each model appear above and to the left; sample traces of each model appear above and to the right (voltage traces in black, input profile in gray, two seconds shown, input frequency 2.5 Hz, total input -3.4 nA/s, as indicated by cyan dot on the false-color image). Total input per second was calculated by integrating input over the entire simulation.

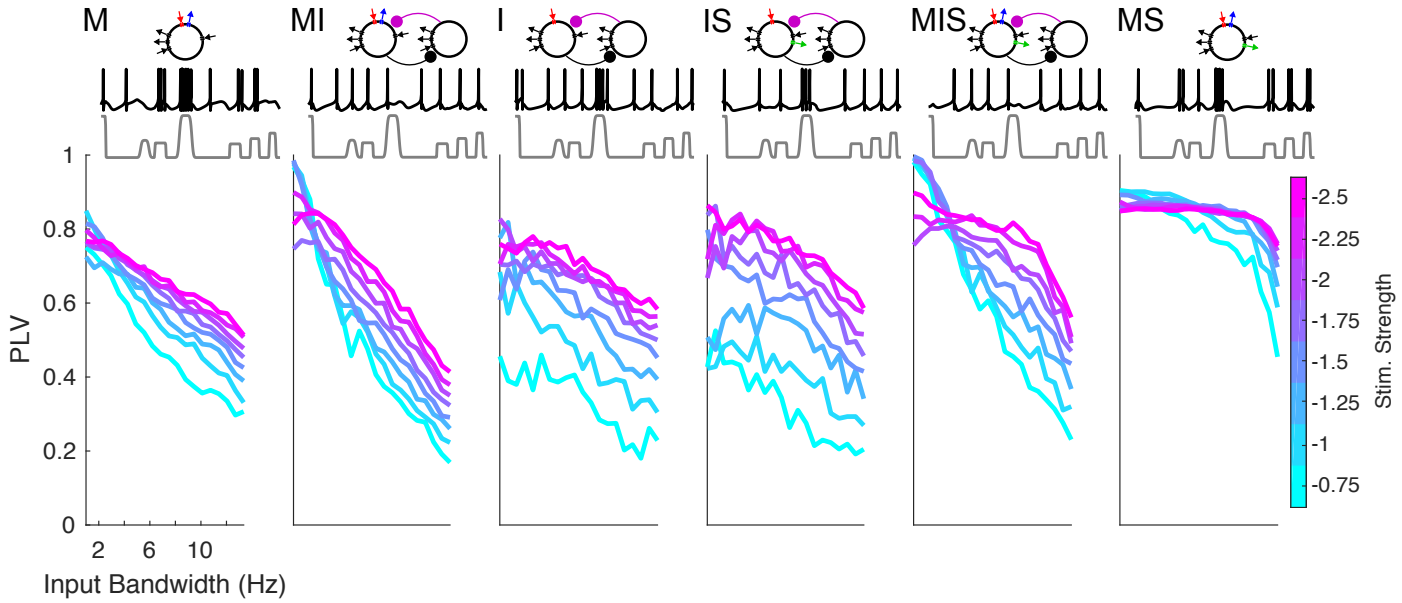

Fig 4. Phase-Locking to Quasi-Rhythmic Inputs. Plots show the (spike-rate adjusted) phase-locking value of spiking to input waveform, for inputs of varying input strength as well as varying bandwidth and regularity (see Methods & Table 5). All inputs have a center frequency of 7 Hz. Schematics of each model appear above. Sample traces from each model are shown in black, in response to inputs shown in gray, having a bandwidth of 10.65 Hz and an input gain of 1.1; 1.1 second total is shown.

2.2.2 Quasi-rhythmic inputs

Next, we tested whether the frequency selectivity of phase-locking exhibited for periodic inputs would carry over to quasi-rhythmic inputs, by exploring the phase-locking behavior of model θ oscillators in response to trains of input pulses in which pulse duration, interpulse duration, and pulse waveform varied from pulse to pulse. The latter were chosen (uniformly) randomly from ranges of pulse "frequencies", "duty cycles", pulse shape parameters, and onset times (see Methods, Equation 3). To create a gradient of sets of (random) inputs with different degrees of regularity, we systematically varied the intervals from which input parameters were chosen (Table 5); we use "bandwidth" here as a shorthand for this multi-dimensional gradient in input regularity. Input pulse trains with a "bandwidth" of 1 Hz were designed to be similar to the 7 Hz periodic pulse trains from Section 2.2.1.

For these "narrowband", highly regular inputs, all six models showed a high degree of phase-locking (Fig. 4). In contrast, phase-locking to "broadband" inputs was high only for the models that exhibited broader frequency ranges of phase-locking to regular rhythmic inputs. At high input strengths, model MS in particular showed a high level of phase-locking that was nearly independent of input regularity (Fig. 4). Notably, model

MIS mirrored the ability of model MS to phase-lock to broadband inputs at high input intensity, while showing frequency selective phase-locking at low input intensity. Indeed, model MIS phase-locked to weak, narrowband quasi-rhythmic inputs better than any other model, perhaps due to its large region of one-to-one phase-locking (Fig. 4).

2.3 Speech entrainment and segmentation

2.3.1 Phase-locking to speech inputs

We tested whether frequency flexibility in response to rhythmic and quasi-rhythmic inputs would translate to an advantage in phase-locking to real speech inputs selected from the TIMIT corpus [64]. We also tested how phase-locking to the speech amplitude envelope might differ between auditory frequency bands, examining the response of each model to 16 different auditory channels, ranging in frequency from 0.1 to 3.3 kHz, extracted, from 20 different sentences selected blindly from the TIMIT corpus, by a model of the cochlea and subcortical nuclei responsible for auditory processing [65]. We varied the input strength of these speech stimuli with a multiplicative gain that varied between 0 and 2, and assessed the PLV of RS cell spiking to auditory channel phase (Fig. 5, Section 4.3). All models exhibited a linear increase in PLV with input gain, and the strongest phase-locking to the mid-vocalic channels ($\sim$0.206-0.411 kHz, with peak phase-locking to 0.357 kHz; $p < 10^{-10}$, Fig. S2). To compare the models' performance without the heterogeneous contribution of sub-optimal channels and gains, we ran further simulations of 1000 sentences at the 0.233 kHz channel (shown to be optimal among a larger number of channels run in the course of our segmentation simulations, see Section 2.3.2 below) and the highest level of multiplicative gain (2). For these simulations, the strength of phase-locking was consistent with the models' ability to phase-lock flexibly to periodic and varied pulse inputs, with the notable exception that models MIS and MI exhibited the weakest performance (Fig. S2), which we hypothesized was again due to their high level of inhibition.

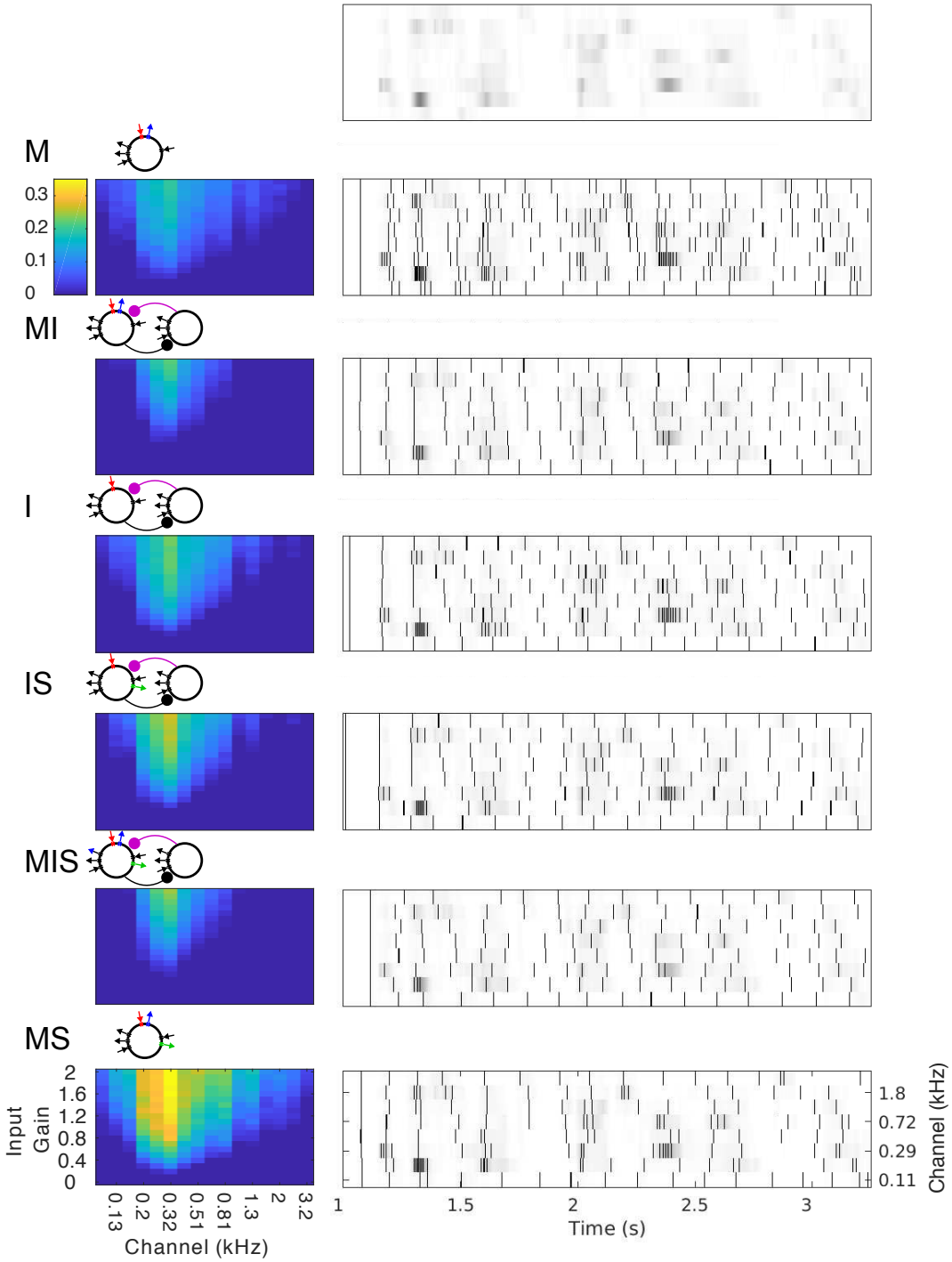

Fig 5. Phase-Locking to Speech Inputs. False-color plots (left) show the mean (spike-rate adjusted) PLV of spiking to speech input waveforms, for different auditory channels (x-axis) as well as varying input strengths (y-axis, see Methods). Gray-scale plots (right) show the spiking response of each model to a selection of 8 auditory channels for a single example sentence. The amplitude of each auditory channel is shown in gray-scale; the top plot shows these amplitudes without any model response. The spiking in response to each channel is overlaid as a raster plot, with a black vertical bar indicating each spike. Schematics of each model appear to the upper left.

2.3.2 Speech segmentation by phase-locked cortical θ oscillators

We next sought to assess whether phase-locking to speech inputs could contribute to functionally relevant speech segmentation, and whether the validity of this segmentation might differ between auditory frequency bands. To do so, we divided the auditory frequency range into 8 sub-bands consisting of 16 channels each, and drove 16 copies of each of our six models with speech input from each sub-band. We used a simple sum-and-threshold mechanism, intended to approximate the integration of the 16 model oscillators' spiking by a shared postsynaptic target, to translate model activity into syllabic-timescale segmental boundaries (see Section 4.4.1). We then compared these model-derived segmental boundaries to transcription-derived boundaries, extracted from phonemic transcriptions of the TIMIT corpus (see Section 4.4.2). Since all our models exhibited the highest levels of phase-locking to the mid-vocalic channels, and since the high energy phase for these channels occurs between syllabic boundaries, we compared model-derived segmental boundaries to the midpoints of transcription-derived syllables, computing a normalized point-process metric $D_{\mathrm{VP},50}$ [66] that penalized model boundaries shifted by more than 50 ms from a syllabic midpoint, as well as "extra" model boundaries and "missed" syllable midpoints (see Section 4.4.3). Because syllabic midpoints are not necessarily linguistically meaningful, the functional utility of model-derived boundaries may not depend on whether they occur exactly at (or within 50 ms of) mid-syllable. Hypothesizing that model-derived boundaries might function simply to identify particular phonemes (i.e., vowels), we also examined the phonemic distribution of model-derived boundaries.

The derivation of boundaries from model spiking depended on two parameters – a decay timescale w_s used to sum spikes over time, and a threshold level r_{thresh} used to determine boundary times. In general, the values of the parameters w_s and r_{thresh} dramatically affected segmentation performance (Fig. S3). Intuitively, these parameters may be thought of as analogous to synaptic timescale and efficacy, or as representing maximal NMDA and AMPA conductances, respectively. The ranking of models' segmentation performance depended on the choice of these parameters (Figure S3), suggesting that a downstream "boundary detector" could "learn" to detect syllable boundaries from the output of the model, by adjusting these parameters.

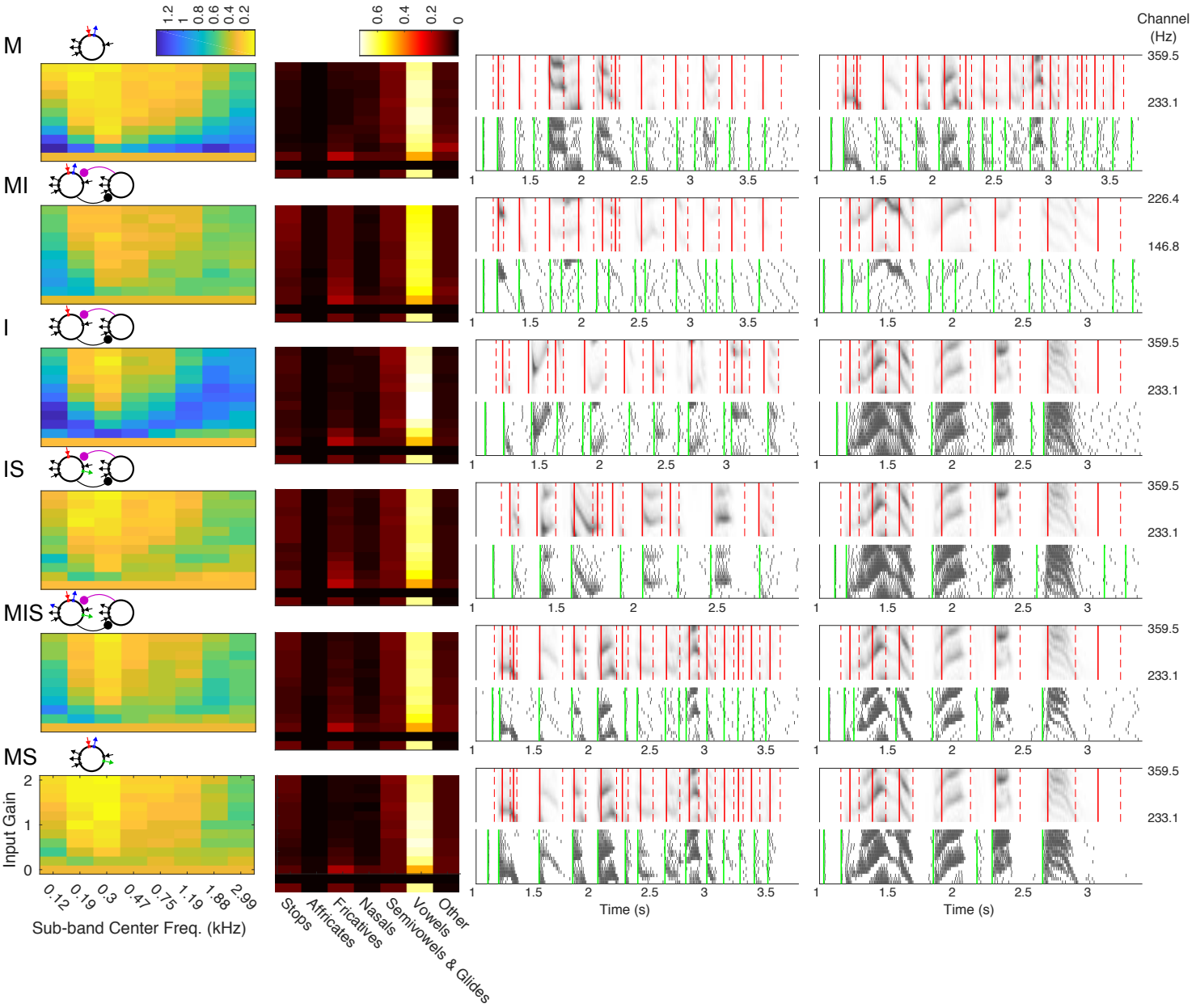

Fig 6. Speech Segmentation. Blue and yellow false-color plots (far left) show mean $D_{\mathrm{VP},50}$ for different auditory sub-bands (x-axis) and varying input strengths (y-axis), for the pair of values taken from $w_s = \{25, 30, \ldots, 75\}$ and $r_{\mathrm{thresh}} = \{1/3, .4, .45, \ldots, .6, 2/3\}$ that minimizes the mean $D_{\mathrm{VP},50}$ for 40 randomly chosen sentences (for any input gain and sub-band). For the same w_s and r_{thresh}, black, red, and white false-color plots (middle left) show the proportion of model-derived boundaries intersecting each phoneme class (x-axis), for the mid-vocalic sub-band (center freq. ∼0.3 kHz) and varying input strengths (y-axis; right). For comparison, the bottom row shows the phoneme distribution of syllable midpoints. Gray-scale plots with green and red lines show example sentences and model responses taken from simulations for the sub-band and input strength with the lowest mean $D_{\mathrm{VP},50}$ for each model: sentences with the lowest (resp. highest) $D_{\mathrm{VP},50}$ for this sub-band and input strength are shown middle right (resp. far right). Each set of two plots shows the speech input (top panel, gray), syllabic boundaries (red dashed lines), and syllable midpoints (red solid lines); as well as the response of the model (bottom, gray) and the model boundaries (green lines). Schematics of each model appear to the upper left.

We thus individually "optimized" each model's performance over a modest set of w_s and r_{thresh} values, finding the w_s and r_{thresh} values for each model that produced the minimum mean $D_{\text{VP},50}$ (for any gain and channel). Comparing $D_{\text{VP},50}$ across these "optimized" data sets (Fig. S4) revealed that segmentation performance roughly mirrored entrainment performance, with model MS, the mid-vocalic sub-band (center frequency 0.296 kHz), and the highest gain producing the lowest mean $D_{\text{VP},50}$.

To more rigorously compare model segmentation performance, we ran simulations with 1000 sentences for only the mid-vocalic channel at the highest gain (2), and once again optimized w_s and r_{thresh} independently for each model (Fig. S4). The resulting ranking across models followed phase-locking flexibility with the exception of model M, which performed as well as model MIS. This tie was surprising, demonstrating the possibility of accurate syllable segmentation even in the absence of high levels of phase-locking to speech inputs. All models, with the exception of model MI, produced a boundary phoneme distribution with a proportion of vowels as high or higher than the proportion of vowels occurring at mid-syllable (Fig. 6).

2.4 Mechanisms of phase-locking

2.4.1 Role of post-input spiking delay

Given that both the most selective and the most flexible oscillator were paced by the m-current, we sought to understand how the dynamics of outward currents contributed to the observed gradient from selective to flexible phase-locking. We hypothesized that phase-locking to input pulse trains in our models depended on the duration of the delay until the next spontaneous spike following a single input pulse. Our rationale was that each input pulse leads to a burst of spiking, which in turn activates the outward currents that pace the models' intrinsic rhythmicity. These inhibitory currents hyperpolarize the models, causing the cessation of spiking for at least a θ period, and in some cases much longer. If the pause in spiking is sufficiently long to delay further spiking until the next input arrives, phase-locking is achieved, given that the next input pulse will also cause spiking (as a consequence of being in the strong forcing regime). In other words, if D is the delay (in s) between the onset of the input pulse and the first

post-input spike, then the lower frequency limit f^* of phase-locking satisfies

$$1/f^* \leq D \Longrightarrow f^* \geq 1/D. \tag{1}$$

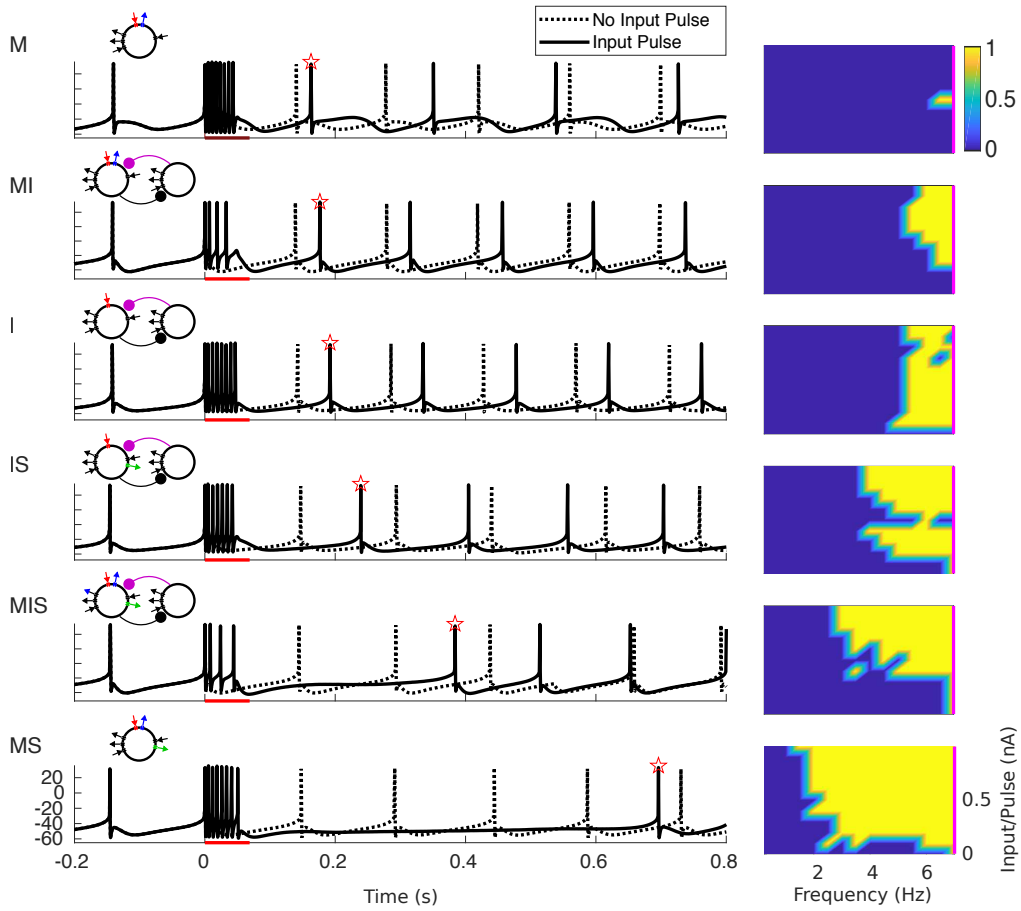

Fig 7. Delay of spiking in response to single pulse determines phase-locking to slow inputs. On the left, voltage traces are plotted for simulations both with (solid lines) and without (dotted lines) an input pulse lasting 50 ms. Red bar indicates the timing of the input pulse; red star indicates the first post-input spike. On the right, the phase-locking value is estimated from the response to a single input pulse using Eqn. 1. Frequency was calculated as 1/(4*(pulse duration)), where pulse duration is in seconds. Input per pulse was calculated by integrating pulse magnitude. The magenta line indicates 7 Hz.

To test this hypothesis, we measured the delay of model spiking in response to single spike-triggered input pulses, identical to single pulses from the periodic inputs discussed in Section 2.2.1, with durations corresponding to periodic input frequencies of 7 Hz or

less, and varied input strengths. The fact that these pulses were triggered by spontaneous rhythmic spiking allowed a comparison between intrinsic spiking and spiking delay post-input (Fig. 7, left), which showed a correspondence between flexible phase-locking and the duration of spiking delay. We also used spiking delay and Eq. 1 to estimate the regions of phase-locking for each model oscillator. In agreement with our hypothesis, the delay-estimated PLV closely matched the profiles of frequency flexibility in phase-locking measured in Section 2.2.1 (Fig. 7, right).

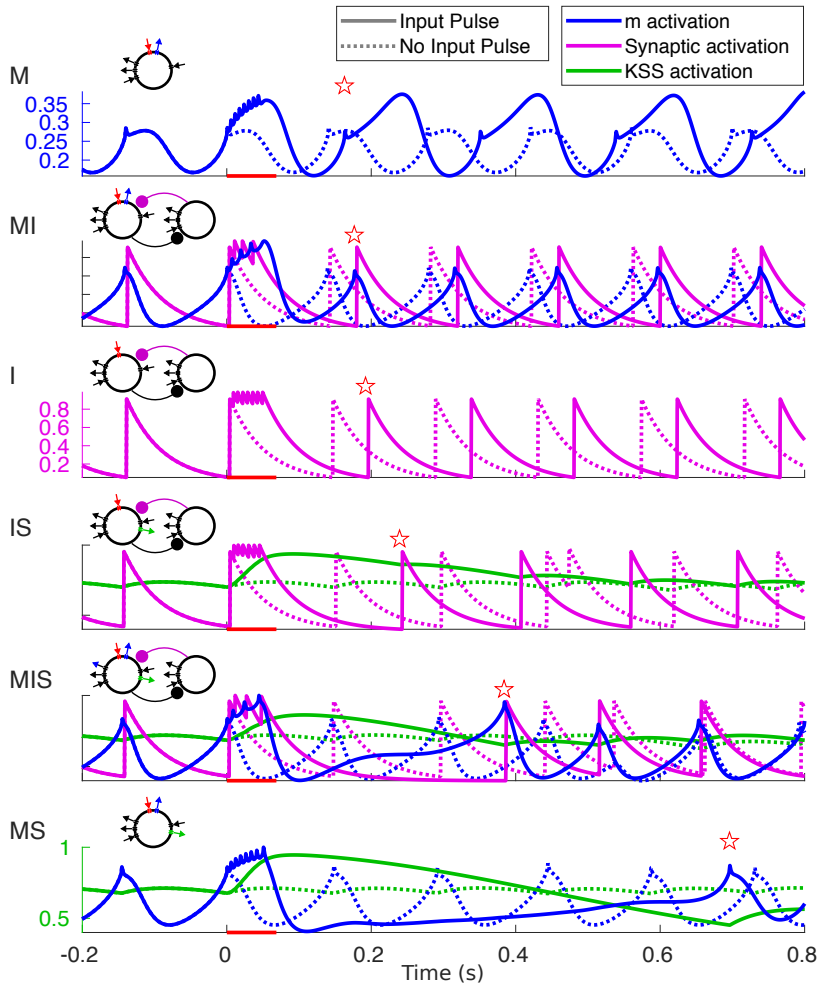

Fig 8. Buildup of outward currents in response to input pulses. Activation variables (color) plotted for simulations both with (dotted lines) and without (solid lines) an input pulse lasting 50 ms. Red bar indicates the timing of the input pulse; red star indicates the time of the first post-input spike.

2.4.2 Dynamics of inhibitory currents

To understand how the dynamics of intrinsic and synaptic currents determined the length of the post-input pause in spiking, we examined the gating variables of the three outward currents simulated in our models during both spontaneous rhythmicity and following forcing with a single input pulse (Fig. 8). Plotting the relationships between these currents during the time step immediately prior to spiking (Fig. 9) offered insights into the observed gradient of phase-locking frequency flexibility. Below, we describe the dynamics of these outward currents, from simple to complex.

Synaptic inhibition Model I spiked whenever the synaptic inhibitory current I_{inh} (Fig. 8, purple) or, equivalently, its gating variable, was sufficiently low. This gating variable decayed exponentially from the time of the most recent SOM cell spike; it did not depend on the level of excitation of the RS cell, and thus did not build up during the input pulse. However, post-input spiking delays did occur because RS and SOM cells spiked for the duration of the input pulse, repeatedly resetting the synaptic inhibitory "clock" – the time until I_{inh} had decayed enough for a spontaneous spike to occur. As soon as spiking stopped (at the end of the input pulse or shortly afterwards – our model SOM interneurons were highly excitable and often exhibited noise-induced spiking after the input pulse), the level of inhibition began to decay, and the next spike occurred one 7 Hz period after the end of the input pulse. For periodic input pulses 1/4 the period of the input rhythm, this suggested that the lower frequency limit f^* of phase-locking for model I was determined roughly by the equation

$$D = \frac{1}{4}\left(\frac{1}{f^*}\right) + \frac{1}{7} \geq \frac{1}{f^*} \quad \Longrightarrow \quad f^* \geq \frac{21}{4} = 5.25,$$

which corresponded to the limit observed for model I in Fig. 3 and Fig. 2.4.1.

m-Current In contrast, model M did not spike when the m-current gating variable reached its nadir, but during the rising phase of its rhythm (Fig. 8). Since the m-current activates slowly, at this phase the upward trajectory in the membrane potential – a delayed effect of the m-current trough – is not yet interrupted by the hyperpolarizing influence of m-current activation. When the cell received an input pulse, the m-current (blue) built up over the course of the input pulse, but since it is a hyperpolarizing current activated by depolarization whose time constant is longest at ∼-26 mV and

shorter at de- or hyperpolarized membrane potentials, this buildup resulted in the m-current rapidly shutting itself off following the input pulse. This rapid drop resulted in a lower trough, and, subsequently, a higher peak value of the m-current's gating variable (because the persistent sodium current had more time to depolarize the membrane potential before the m-current was activated enough to hyperpolarize it), changing the frequency of subsequent STOs. It didn't, however, affect the model's phase-locking in the strong forcing regime; the fast falling phase of the m-current following the pulse kept the post-input delay small (Fig. 8). This "elastic" dynamics offers an explanation for model M's inflexibility: the buildup of m-current during an input pulse leads to a fast hyperpolarization of the membrane potential, which, in turn, causes rapid deactivation of the m-current and subsequent rapid "rebound" of the membrane potential to depolarized levels, preserving the time of the next spike.

Super-slow K current In models with a super-slow K current, this current, like synaptic inhibition, decayed to a nadir before each spike of the intrinsic rhythm. Unlike synaptic inhibition, $I_{\mathrm{K_{SS}}}$ activation built up dramatically during an input pulse (Fig. 8, green), and decayed slowly, increasing the latency of the first spike following the input pulse substantially (Fig. 2.4.1). This slow-building outward current interacted differently, however, with synaptic and intrinsic θ-timescale currents. In model IS, both I_{inh} and $I_{\mathrm{K_{SS}}}$ decayed monotonically following an input pulse, until the total level of hyperpolarization was low enough to permit another spike. We hypothesized that $I_{\mathrm{K_{SS}}}$ and I_{inh} interacted additively to produce hyperpolarization and a pause in RS cell spiking. In other words, the delay until the next spike was determined by the time it took for a sum of the two currents' gating variables (weighted by their conductances and the driving force of potassium) to drop to a particular level. The fact that we expect this weighted sum of the gating variables to be nearly the same (having value, say, a^*) in the time t^* before each spike suggests that the two gating variables are negatively linearly related at spike times:

$$
\begin{aligned}
g_{\mathrm{SOM\to RS}} s(t^*)(V(t^*) - E_{\mathrm{K}}) + g_{\mathrm{K_{SS}}} q(t^*)(V(t^*) - E_{\mathrm{K}}) &\simeq a^* \\
\Longrightarrow q(t^*) \simeq -\frac{g_{\mathrm{SOM\to RS}}}{g_{\mathrm{K_{SS}}}} s(t^*) &+ \frac{a^*}{g_{\mathrm{K_{SS}}}(V(t^*) - E_{\mathrm{K}})}.
\end{aligned}
$$

Plotting the activation levels of these two currents in the timestep before each spike

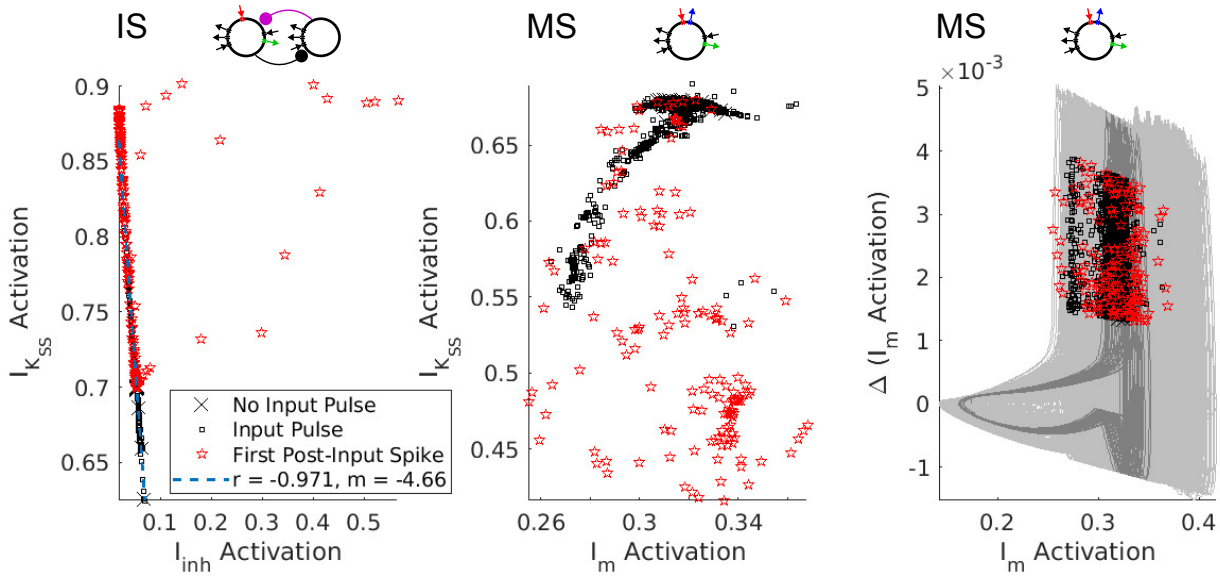

Fig 9. Linear vs. Synergistic Interactions of Inhibitory Currents. Plots of the pre-spike gating variables in models IS and MS. Left, the pre-spike activation levels of I_{inh} and $I_{\mathrm{K_{SS}}}$ in model IS have a negative linear relationship. (Regression line calculated excluding points with I_{inh} activation > 0.1.) Middle, the pre-spike activation levels of I_{m} and $I_{\mathrm{K_{SS}}}$ in model MS do not exhibit a linear relationship. Right, the plotting the activation level of I_{m} against its first difference reveals that pre-spike activation levels are clustered along a single branch of the oscillator's trajectory. (Light gray curves represent trajectories with an input pulse; dark gray curves represent trajectories without an input pulse.

(excluding forced spikes and a handful of outliers) against each other confirmed this hypothesis (Fig. 9, model IS).

The interaction between I_{m} and $I_{\mathrm{K_{SS}}}$ was more complex, as seen in model MS. The pre-spike activation levels of these two currents were not linearly related (Fig. 9, center). When $I_{\mathrm{K_{SS}}}$ built up, it dramatically suppressed the level of the m-current gating variable, biasing the competition between I_{m} and $I_{\mathrm{Na_P}}$ and reducing STO amplitude, and the $I_{\mathrm{K_{SS}}}$ activation had to decay to levels much lower than "baseline" before the oscillator would spike again. Indeed, spiking appeared to require m-current activation to return *above* "baseline", and also to be in the rising phase of its oscillatory dynamics. The dependence of spiking on the phase of the m-current activation could be seen by plotting the "phase plane" trajectories of the oscillator – plotting the m activation against its first difference immediately prior to each spike – revealing a branch of the oscillator's periodic trajectory along which pre-spike activation levels were clustered (Fig. 9, right). Plotting the second difference against the first revealed similar periodic dynamics (Fig. S5, top row).

The models containing both synaptic inhibition and m-current exhibited similar dynamics to model MS, with a dependence of spiking on the phase of the rhythm in I_{m} activation being the clearest pattern observable in the pre-spike activation variables (Fig. S5). This suggests that the delay following the input pulse in these models also reflects an influence of θ-timescale STOs, which may exhibit more complex interactions with I_{inh} in model MI, similar qualitatively if not quantitatively to their interactions with $I_{\mathrm{K_{SS}}}$ in models MS and MIS.

3 Discussion

Our results link the biophysics of cortical oscillators to speech segmentation via flexible phase-locking, suggesting that the intrinsic inhibitory currents observed in cortical θ oscillators [49, 50] may enable these oscillators to entrain robustly to θ-timescale fluctuations in the speech amplitude envelope, and that this entrainment may provide a substrate for enhanced speech segmentation that reliably identifies mid-syllabic vocalic nuclei. They trace the capacity of cortical θ oscillators for flexible phase-locking to synergistic interactions between their intrinsic currents, and demonstrate that similar oscillators lacking either of these intrinsic currents show markedly less frequency flexibility in phase-locking, regardless of the presence of θ-timescale synaptic inhibition. These findings suggest that synaptic and intrinsic inhibition may tune neural oscillators to exhibit different levels of phase-locking flexibility, allowing them to play diverse roles – from reliable internal clocks to flexible parsers of sensory input – and have consequences for neural dynamics, speech perception, and brain function.

3.1 Mechanisms of phase-locking

For models containing a variety of intrinsic and synaptic currents, spiking delay following a single input pulse was an important determinant of the lower frequency limit of phase-locking in the strong-forcing regime (Fig. 7). A super-slow current, $I_{\mathrm{K_{SS}}}$, aided the ability to phase-lock to slow frequencies in our models, by building up over a slow timescale in response to burst spiking during a long and strong input pulse. The presence of the super-slow K current increased the frequency range of phase-locking, with every model containing $I_{\mathrm{K_{SS}}}$ able to phase-lock to slower periodic inputs than any

model without $I_{\mathrm{K_{SS}}}$ (Fig. 3). The fixed delay time of synaptic inhibition seemed to stabilize the frequency range of phase-locking, while the voltage-dependent and "elastic" dynamics of the m-current seemed to do the opposite. Specifically, the four models containing I_{inh} exhibited an intermediate frequency range of phase-locking, while both the narrowest and the broadest frequency ranges of phase-locking occurred in the four model θ oscillators containing I_{m}; and the very narrowest and broadest ranges occurred in the two models containing I_{m} and lacking I_{inh} (Fig. 3).

Our investigations showed that the flexible phase-locking in models MS and MIS resulted from a synergistic interaction between slow and super-slow K currents, demonstrated here – to our knowledge – for the first time. We conjecture that this synergy depends on the subthreshold oscillations (STOs) engendered by the slow K current (the m-current) in our models, as was suggested by an analysis of the pre-spike activation levels of the inhibitory currents in models IS and MS. In model IS, there were no STOs, and the interaction between θ-timescale inhibition (which was synaptic) and $I_{\mathrm{K_{SS}}}$ was additive, so that spikes occurred whenever the (weighted) sum of these gating variables dropped low enough (Fig. 9). In models MIS and MS, where STOs resulted from interactions between the m-current and the persistent sodium current, spiking depended not only on the level of activation of the m-current, but also on the phase of the endogenous oscillation in m-current activation (Fig. 9).

For all models, the frequency flexibility of phase-locking to periodic inputs translated to the ability to phase-lock to quasi-rhythmic (Fig. 4) and speech (Fig. 5) inputs. While it is reasonable to hypothesize that this is the result of the mechanism of phase-locking in the regime of strong forcing, it is important to note that imperfect phase-locking in our models resulted not only from "extra" spikes in the absence of input (as predicted by this hypothesis), but also from "missed" spikes in the presence of input (Fig. 4). A dynamical understanding of these "missed" spikes may depend on the properties of our oscillators in the weak-forcing regime.

Phase-locking of neural oscillators under weak forcing has been studied extensively [54–58]. In this regime, a neural oscillator stays close to a limit cycle during and after forcing, and as a result the phase of the oscillator is well-defined throughout forcing. Furthermore, the change in phase induced by an input is small (less than a full cycle), can be calculated, and can be plotted as a function of the phase at which the

input is applied, resulting in a phase-response curve (PRC). Our results pertain to a dynamical regime in which PRC theory does not apply, since our forcing is strong and long enough that our oscillators complete multiple cycles during the input pulse, and as a result the phase at the end of forcing is not guaranteed to be a function of the phase at which forcing began. Furthermore, in oscillators which contain $I_{\mathrm{K_{SS}}}$, the dynamics of this slow current add an additional dimension, which makes it impossible to describe the state of these oscillators in terms of a simple phase variable. Not only the phase of the oscillator, but also its amplitude (which is impacted by the activation of $I_{\mathrm{K_{SS}}}$), determine its dynamics.

Previous work has illuminated many of the dynamical properties of the θ-timescale m-current. The addition of an m-current (or any slow resonating current, such as an h-current or other slow non-inactivating K current) changes a neuron from a Type I to a Type II oscillator [67, 68]. The generation of membrane potential resonance (and subthreshold oscillations) by resonating currents is well-studied [49, 69, 70], and recently it has been shown that the θ-timescale properties of the M-current allow an E-I network subject to θ forcing to precisely coordinate with external forcing on a γ timescale [61]. While STOs play an important role in the behaviors of our model oscillators, subthreshold resonance does not automatically imply suprathreshold resonance or precise response spiking [71]. Thus, our results are not predictable (either *a priori* or *a posteriori*) from known effects of the m-current on neuronal dynamics.

Larger (synaptic) inhibition-paced networks have been studied both computationally and experimentally [52, 72–75], and can exhibit properties distinct from our single (RS) cell inhibition-paced models: computational modeling has shown that the addition of E-E and I-I connectivity in E-I networks can yield frequency flexibility through potentiation of these recurrent connections [73, 75]; and experimental results show that amplitude and instantaneous frequency are related in hippocampal networks, since firing by a larger proportion of excitatory pyramidal cells recruits a larger population of inhibitory interneurons [74], a phenomenon which may enable more frequency flexibility in phase-locking. This raises the question of why the brain would select phase-locking flexibility in single cells vs. networks. One possible answer is energetic efficiency. If flexibility in an inhibition-paced oscillatory network depends on recruiting large numbers of inhibitory interneurons, it may be more efficient to utilize a small number of

oscillators, each capable (on its own) of entrainment to quasi-rhythmic inputs containing a large range of instantaneous frequencies.

3.2 Functional implications for neuronal entrainment to auditory and speech stimuli

Our focus on the θ timescale is motivated by results underscoring the prominence of theta rhythms in the spontaneous and stimulus-driven activity of primate auditory cortex [43, 76–78] and by evidence for the (causal [39–42]) role of δ/θ frequency speech-brain entrainment in speech perception [32–39, 42]. Our results suggest that the types of inhibitory currents pacing cortical θ oscillators with an intrinsic frequency of 7 Hz determine these oscillators' ability to phase-lock to the (subcortically processed [65]) amplitude envelopes of continuous speech. While an oscillator with an intrinsic frequency of 3 Hz might do an equally good job of phase-locking to strong inputs with frequencies between 3 and 9 Hz, this does not seem to be the strategy employed by the auditory cortex: the frequencies of (low-frequency) oscillations in primate auditory cortex are $\sim$1.5 and $\sim$7 Hz, not 3 Hz [43]; existing experimental [43, 79] and computational [80] evidence suggests that cortical δ oscillators are unlikely to be driven at θ frequencies even by strong inputs; and MEG studies show that across individuals, speech comprehension is high when cortical frequencies are the same as, or higher than, speech envelope frequencies, and becomes poorer as this relationship reverses [81].

Another important question raised by our results (and by one of our reviewers) is the following: If flexible entrainment to a (quasi-)periodic input depends on the lengths of the delays induced by the input, why go to the trouble of using an oscillator at all, rather than a cell responding only to sufficiently strong inputs? The major difference between oscillators and non-oscillatory circuits driven by rhythmic inputs is what happens when the inputs cease (or are masked by noise): while a non-oscillatory circuit lapses into quiescence, an oscillator continues spiking at its endogenous frequency. Thus, oscillatory mechanisms can track the temporal structure of speech through interruptions and omissions in the speech signal [16]. This capability is crucial to the adjustment of speech processing to the speech rate, a phenomenon in which brain oscillations are strongly implicated. While (limited) speeding or slowing of *entire* utterances does not

affect their intelligibility, altering context speech rate can change the perception of unaltered target words, even making them disappear [82–87]. In recent MEG experiments, brain oscillations entrained to the rhythm of contextual speech persisted for several cycles after a speech rate change [87], with this sustained rhythmic activity associated with altered perception of vowel duration and word identity following the rate change [87]. Multiple hypothetical mechanisms have been proposed to account for these effects: the syllabic rate (as encoded by the frequency of an entrained θ rhythm) may determine the sampling rate of phonemic fine structure (as effected by γ rhythmic circuits) [6, 53]; predictive processing of speech may use segment duration relative to context speech speed as evidence to evaluate multiple candidate speech interpretations [47]; and oscillatory entrainment to the syllabic rate may time relevant calculations, enabling the optimal balance of speed and accuracy in the passing of linguistic information up the processing hierarchy before the arrival of new input – so-called "chunk-and-pass" processing [88].

Recent experiments shed light on the limits of adaptation to (uniform) speech compression, showing that while cortical speech-brain phase entrainment persisted for syllabic rates as high as 13 Hz (a speed at which speech was not intelligible), β-rhythmic activity was abnormal in response to this unintelligible compressed speech [89]. This work suggests that the upper syllabic rate limit on speech intelligibility arises not from defective phase-locking, but from inadequate time for mnemonic or other downstream processes between syllables [89]. This agrees with our finding that the upper frequency boundary on phase-locking extends well above the upper syllabic rate boundary on speech intelligibility ($\sim$9 Hz), and is largely determined by input strength. Nonetheless, it is noteworthy that task-related auditory cortical entrainment operates most reliably over the 1-9 Hz (syllabic) ranges [76]. Further exploration of how speech compression affects speech entrainment by neuronal oscillators is called for.

Out of our models, MS came closest to spiking selectively at the peaks of the speech amplitude envelope, yet it did not perform perfectly. This was to be expected for a signal as broadband and irregular as the amplitude envelope of speech, which presents challenges to both entrainment and its measurement (see Section 4.3.3). As we've mentioned, defects in phase-locking were also due to both "missed" cycles and "extra" spikes (Fig. 5). The frequency of occurrence of these missed and extra spikes was

traded off as tonic excitation to model MS was varied: lower levels of tonic excitation led to more precise phase-locking (i.e., fewer extra spikes) but more missed cycles, and higher levels of tonic excitation lead to less precise phase-locking but a lower probability of any cycle being missed (Fig. S5).

3.3 Functional implications for speech segmentation

Multiple theories suggest a functional role for cortical θ oscillations in segmenting auditory and speech input at the syllabic timescale [6, 11–13, 16, 39, 42, 78, 90, 91]. To explore whether the different levels of phase-locking to speech inputs observed in our models might have consequences for syllabic segmentation, we implemented a simple method to extract putative segmental boundaries from the spiking of multiple (unconnected) copies of our models. Our results serve to demonstrate that the accuracy with which segmental boundaries can be extracted from the spiking of speech-entrained cortical oscillators depends on the particular biophysics of those oscillators. They suggest that the information in the mid-vocalic channels provides an advantage for entrainment to speech and for syllabic-timescale segmentation. Finally, they open the door to many new questions about the neuronal bases of speech processing.

Our work points to frequency flexibility, which appears to enable segmentation accuracy even at low levels of entrainment to the speech signal (as can be seen by contrasting the segmentation performance of models MIS and MI), as one of the factors that can impact segmentation accuracy. However, it is clear that other factors also contribute. One likely factor is excitability, which has been a minor theme, contributing second-order effects to the behaviors of models MI, MIS, and MS (Fig. S1, S6). While we tuned our models to exhibit the same (7 Hz) frequency of tonic spiking in the absence of (dynamic) input, and attempted to qualitatively match F-I their curves, our models exhibited clear differences in the number of spikes evoked by inputs of the same strength (Fig. 5, 7). It is likely that this in turn impacted the sum-and-threshold mechanism used to extract syllable boundaries. A highly excitable oscillator may respond to speech input with a surfeit of spiking from which accurate syllable boundaries can be carved by the choice of w_s and r_{thresh}; such a mechanism may account for the unexpectedly accurate segmentation performance of model M. The issue

of excitability arises again when inquiring into the advantages mid-vocalic channels offer for speech entrainment and segmentation, as these channels differ not only in their frequency content but also in their spectral power, having higher amplitude than other channels. We have chosen not to normalize speech input beyond the transformations implemented by a model of subcortical auditory processing, but investigating how different types of normalization affect speech entrainment and segmentation could illuminate whether mid-vocalic channels' frequency, amplitude, or both are responsible for the heightened functionality they drive.

There remains much to explore about how segmental boundaries may be derived from the spiking of populations of cortical oscillators. While our implementation was extremely simplistic, omitting heterogeneity in parameters or synaptic or electrical connectivity between oscillators, "optimized" model-derived boundaries seemed to arise from a relatively complex integration of the rich temporal dynamics of population activity within a given sub-band (Fig. 6). This is somewhat in contrast to the regular and highly synchronous spike volleys characterizing previous models of oscillatory syllable segmentation, in which all θ oscillators received the same channel-averaged speech input [45]. In our implementation, a boundary is signaled when the activity of the oscillator network passes a given threshold. This is in agreement with recent results showing that neurons in middle STG, a region of auditory cortex implicated in syllable and word recognition, respond to acoustic onset edges (i.e., peaks in the rate of change of the speech amplitude envelope) [92, 93]. It may also explain why segmentation failures occurred when the speech amplitude envelope remained high through an extended time period that included multiple syllabic boundaries (Fig. 6, rightmost panels). One way around this is to combine information across, as well as within, auditory sub-bands. While our work supports the hypothesis that identification of vocalic nuclei, rather than consonantal clusters, is associated with more precise syllabic-timescale segmentation, it doesn't preclude the use of information about the timing of consonantal clusters to aid segmentation. Interestingly, auditory cortical regions encoding different auditory frequencies entrained to different phases of rhythmic stimuli presented at 1.6 Hz – with 11-15 kHz regions firing during high-amplitude phases, and all other regions firing in antiphase – an alternating response pattern suggested to relate to the alternation of vowels and consonants in speech [94]. We suggest that a deeper understanding of the

dynamic repertoire afforded by the simple model presented here may provide a foundation for future investigations of more complex (and realistic) networks.

Prevous work showed that θ bursts generated by an oscillator paced by synaptic inhibition are able to predict syllable boundaries "on-line" at least as accurately as state-of-the-art offline syllable detection algorithms [45]. While we have not compared our models directly to these syllable detection algorithms, we explore the performance of synaptic inhibition-paced θ oscillators whose mechanisms of entrainment are similar to those modeled in previous work. In our simulations, models paced even in part by synaptic inhibition perform uniformly worse at syllabic-timescale segmentation than similar models paced by intrinsic currents alone. However, there exist several differences between previous and current models – the source of drive (channel averaged and passed through an optimized spectrotemporal filter vs. frequency specific "raw" input), models at different levels of complexity (leaky integrate-and-fire as opposed to Hodgkin-Huxley in this work), different temporal dynamics of synaptic inhibition (a longer rise time of synaptic inhibition in earlier models), and more rigorous optimization of parameters for a single oscillatory model – all of which may lead to differences in performance between model implementations.

This earlier work positioned syllable segmentation and speech recognition by oscillatory networks within the landscape of syllable detection algorithms arising from the fields of linguistics, engineering, and artificial intelligence [45]. While the current work has focused more on how the biophysics of neuronal oscillators impact the implementation of speech entrainment and segmentation, an understanding of how differences in segmentation performance and location affect speech recognition is an important direction for future work. It remains unclear whether the explicit representation of segmental boundaries contributes to the effects of speech rate and oscillatory phase on syllable and word recognition [62, 78, 82–87], or to the proposed underlying mechanisms that implicate speech segmentation at the neuronal level [6, 47, 53, 88]. Indeed, whether speech recognition in general requires explicit segmentation or only the entrainment of cortical activity to the speech rhythm remains obscure. Cortical θ oscillators are embedded in a stimulus-entrainable cortical rhythmic hierarchy [43, 91, 94–96], receiving inputs from deep IB cells embedded in δ-rhythmic circuits [43, 50, 63, 96], and connected via reciprocal excitation to superficial RS cells

embedded in β- and γ-rhythmic circuits [50, 80]. In the influential TEMPO framework, the θ oscillator is hypothesized to be driven by δ circuits, and to drive γ circuits, with a linkage between θ and γ frequency adjusting the sampling rate of auditory input to the syllabic rate [6, 53]. It has been hypothesized that cortically-identified syllabic boundaries may reset the activity of γ-rhythmic circuits responsible for sampling and processing incoming syllables, a reset necessary for accurate syllable recognition [6, 44, 47, 53]. By indicating the completion of the previous syllabic segment, they may also trigger the activity of circuits responsible for updating the linguistic interpretations of previous speech [53]. Not only this reset cue, but also θ-rhythmic drive to γ-rhythmic circuits, is necessary for accurate syllable decoding within this framework [45]. Recent work with leaky-integrate-and-fire models demonstrates that top-down spectro-temporal predictions can be integrated with theta-gamma coupling, with the latter enabling the temporal alignment of the latter acoustic input [47].

Using the output of our models as an input to syllable recognition circuitry – perhaps via γ-rhythmic circuits [44, 45, 47] – would enable exploration of whether the differences in segmentation accuracy we uncover are functionally relevant for speech recognition. Comparing syllable recognition when these circuits are driven by model-derived segmental boundaries vs. when they are driven only by model spiking may shed light on whether the explicit representation of segmental boundaries is necessary to syllable recognition. Such research would also provide an opportunity to test claims that "theta syllables" provide more information for syllabic decoding than conventional syllables [16]. Our results support the hypothesis that cortical θ oscillators align with speech segments bracketed by vocalic nuclei – so-called "theta syllables" – as opposed to conventional syllables, which defy attempts at a consistent acoustic characterization, but are (usually) bracketed by consonantal clusters [16]. These "theta-syllables" are suggested to have information-theoretic advantages over conventional linguistic syllables: the vocalic nuclei of speech have relatively large amplitudes and durations, making them prominent in noise and reliably identifiable [19]; and windows whose edges align with vocalic nuclei center the diphones that contain the majority of the information for speech decoding, ensuring this information is sampled with high fidelity. These claims, if they prove to have functional relevance, may illuminate how speech-brain entrainment aids speech comprehension in noisy or

otherwise challenging environments [97–99]. Connecting the complex and rich dynamics of networks of biophysically detailed neuronal oscillators to plausible speech recognition circuitry may uncover novel functional and mechanistic factors contributing to speech processing and its dysfunctions [100–104].

3.4 Versatility in cortical processing through flexible and restricted entrainment

More broadly, there is evidence that cortical θ oscillators in multiple brain regions, entrained to distinct features of auditory and speech inputs, may implement a variety of functions in speech processing. Different regions of human superior temporal gyrus (STG) respond differentially to speech acoustics: posterior STG responds to the onset of speech from silence; middle STG responds to acoustic onset edges; and anterior STG responds to ongoing speech [92, 93]. Similarly, bilaterally occuring δ/θ speech-brain entrainment may subserve hemispherically distinct but timescale-specific functions, with right-hemispheric phase entrainment [96] encoding acoustic, phonological, and prosodic information [33, 96, 98, 105, 106], and left-hemispheric amplitude entrainment [96] encoding higher-level speech structure [38, 107–109] and top-down predictions [110–112]. Frequency flexibility may shed light on how these multiple θ oscillations are distinguished, collated, and combined. One tempting hypothesis is that the gradient from flexible to restricted phase-locking corresponds to a gradient from stimulus-entrained to endogenous brain rhythms, with oscillators closer to the sensory periphery exhibiting more flexibility and reverting to intrinsic rhythmicity in the absence of sensory input, enabling them to continue to couple with central oscillators that exhibit less phase-locking flexibility. It is suggestive that the conductance of the m-current, which is key to flexible phase-locking in our models, is altered by acetylcholine, a neuromodulator believed to affect, generally speaking, the balance of dominance between modes of internally and externally generated information [63, 113–115].

Even the potential for flexible entrainment, however, does not seem to be ubiquitous in the brain. In the hippocampus, for example, the θ rhythm is robustly periodic, exhibiting relatively small frequency changes with navigation speed [116]. Interestingly,

the mechanisms of hippocampal θ and the neocortical θ rhythmicity discussed in this paper are very different: while the former is dominated by synaptic inhibition, resulting from an interaction of synaptic inhibition and the h-current in oriens lacunosum moleculare interneurons [48], the latter is only modified by it [50]. Our results suggest that mechanisms like that of hippocampal θ, far too inflexible to perform the segmentation tasks necessary for speech comprehension, are instead optimized for a different functional role. One possibility is that imposing a more rigid temporal structure on population activity may help to sort "signal" from "noise" - i.e., imposing a strict frequency and phase criterion that inputs must meet to be processed, functioning as a type of internal clock. Another possibility is that more rigidly patterned oscillations result from a tight relationship to motor sampling routines which operate over an inherently more constrained frequency range, as, for example, whisking, sniffing, and running are related to hippocampal θ [117, 118].

Along these lines, it is intriguing that model MIS exhibits both frequency selectivity in phase-locking at low input strengths, and frequency flexibility in phase-locking at high input strengths (Fig. 4). Physiologically, input gain can depend on a variety of factors, including attention, stimulus novelty and salience, and whether the input is within- or cross-modality. A mechanism that allows input gain to determine the degree of phase-locking frequency flexibility could enable the differential processing of inputs based on these attributes. It is tempting to speculate that such differential entrainment may play a role in both the low levels of speech entrainment of model MIS, and in the model's ability to carry out accurate segmentation in spite of it. Perhaps more trenchantly, our paper shows that flexible phase-locking requires the activation of intrinsic outward currents which are themselves responsive to multiple neuromodulators (the m-current, for example, is decreased by muscarinic acetylcholine receptor activation). Thus, the phase-locking properties of neurons are themselves modulable, allowing the same neurons to entrain differently to rhythmic inputs depending on the neuromodulatory context.

Although from one perspective model MIS is the most physiologically realistic of our models, as neurons in deep cortical layers are likely to exhibit all three outward currents studied in this paper [50], the minimal impact of synaptic inhibition on these large pyramidal cells suggests that model MS is a functionally accurate representation of the

majority (by number) of RS cells in layer 5 [63]. It thus represents the main source of θ rhythmicity in primary neocortex [63], and a major source of cortico-cortical afferents driving "downstream" processing [119, 120]. Its properties may have strong implications for the biophysical mechanisms used by the brain to adaptively segment and process complex auditory stimuli evolving on multiple timescales, including speech.

4 Methods

All simulations were run on the MATLAB-based programming platform DynaSim [121], a framework specifically designed by our lab for efficiently prototyping, running, and analyzing simulations of large systems of coupled ordinary differential equations, enabling in particular evaluation of their dynamics over large regions of parameter space. DynaSim is open-source and all models will be made publicly available using this platform.

4.1 Model equations

Our models consisted of at most two cells, a regular spiking (RS) pyramidal cell and an inhibitory interneuron with a timescale of inhibition like that observed in somatostatin-positive interneurons (SOM). Each cell was modeled as a single compartment with Hodgkin-Huxley dynamics. In our RS model, the membrane currents consisted of fast sodium (I_{Na}), delayed-rectifier potassium ($I_{\mathrm{K_{DR}}}$), leak (I_{leak}), slow potassium or m- (I_{m}), and persistent sodium ($I_{\mathrm{Na_P}}$) currents taken from a model of a guinea-pig cortical neuron [49], and calcium (I_{Ca}) and super-slow potassium ($I_{\mathrm{K_{SS}}}$, calcium-activated potassium in this case) currents with dynamics from a hippocampal model [122]. The voltage $V(t)$ was given by the equation

$$C\frac{dV}{dt} = I_{\mathrm{app}} - I_{\mathrm{Na}} - I_{\mathrm{K_{DR}}} - I_{\mathrm{leak}} - I_{\mathrm{m}} - I_{\mathrm{Na_P}} - I_{\mathrm{Ca}} - I_{\mathrm{K_{SS}}} - I_{\mathrm{inh}}$$

where the capacitance $C = 2.7$ reflected the large size of deep-layer cortical pyramidal cells, and I_{app}, the applied current, was given by

$$I_{\mathrm{app}}(t) = g_{\mathrm{app}}\left[\left(\frac{t}{\tau_{\mathrm{trans}}}\chi_{\{t\leq\tau_{\mathrm{trans}}\}}(t) + \chi_{\{t>\tau_{\mathrm{trans}}\}}(t)\right) + p_{\mathrm{noise}}W(t)\right]$$

Table 1. Currents.

I_{Na} & $I_{\mathrm{Na,SOM}}$	$g_{\mathrm{Na}} m_{\mathrm{Na}}^3 h\,(V - E_{\mathrm{Na}})$
$I_{\mathrm{K_{DR}}}$ & $I_{\mathrm{K_{DR},SOM}}$	$g_{\mathrm{K_{DR}}} m_{\mathrm{K_{DR}}}^4\,(V - E_{\mathrm{K}})$
I_{leak} & $I_{\mathrm{leak,SOM}}$	$g_{\mathrm{leak}}\,(V - E_{\mathrm{leak}})$
I_{m}	$g_{\mathrm{m}} n\,(V - E_{\mathrm{K}})$
$I_{\mathrm{Na_P}}$	$g_{\mathrm{Na_P}} m_{\mathrm{Na_P}}\,(V - E_{\mathrm{Na_P}})$
I_{Ca}	$g_{\mathrm{Ca}} s^2\,(V - E_{\mathrm{Ca}})$
$I_{\mathrm{K_{SS}}}$	$g_{\mathrm{K_{SS}}} q\,(V - E_{\mathrm{K}})$
I_{inh} & I_{exc}	$g_{\mathrm{pre\to post}} s_{\mathrm{pre\to post}}\,(V_{\mathrm{post}} - E_{\mathrm{pre\to post}})$

Table 2. Equilibrium Voltages.

	RS	FS
E_{Na}	40	50
E_{K}	-80	-95
E_{leak}	-65	-70
$E_{\mathrm{Na_P}}$	50	–
E_{Ca}	120	–
$E_{\mathrm{RS\to SOM}}$		0
$E_{\mathrm{SOM\to RS}}$	-95	

where $\chi_{\mathrm{S}}(t)$ is the function that is 1 on set S and 0 otherwise, the transition time $\tau_{\mathrm{trans}} = 500$ ms, the noise proportion $p_{\mathrm{noise}} = 0.25$, and $W(t)$ a white noise process. (The applied current ramps up from zero during the first 500 ms to minimize the transients that result from a step current.) For SOM cells, the membrane currents consisted of fast sodium ($I_{\mathrm{Na,SOM}}$), delayed-rectifier potassium ($I_{\mathrm{K_{DR},SOM}}$), and leak ($I_{\mathrm{leak,SOM}}$) currents [123]. The voltage $V(t)$ was given by the equation

$$C_{\mathrm{SOM}} \frac{dV}{dt} = I_{\mathrm{app,SOM}} - I_{\mathrm{Na,SOM}} - I_{\mathrm{K_{DR},SOM}} - I_{\mathrm{leak,SOM}} - I_{\mathrm{exc}}$$

where $C_{\mathrm{SOM}} = 0.9$ and $I_{\mathrm{app,SOM}}$, the applied current, is constant in time. The form of each current is given in Table 1; equilibrium voltages are given in Table 2; and conductance values for all six models that are be introduced in *Results: Modeling cortical θ oscillators* (see Figure 1) are given in Table 3.

The dynamics of activation variable x (ranging over h, $m_{\mathrm{K_{DR}}}$, n, $m_{\mathrm{Na_P}}$, s, and q in Table 1) were given either in terms of its steady-state value x_∞ and time constant τ_x by the equation

$$\frac{dx}{dt} = \frac{x_\infty - x}{\tau_x},$$

Table 3. Maximal Conductances.

Model	M	MI	I	IS	MIS	MS
g_{Na}	125	125	125	125	125	125
$g_{\mathrm{K_{DR}}}$	54	54	54	54	54	54
g_{leak}	0.31	0.27	0.78	0.78	0.27	0.27
g_{m}	1.4472	1.4472	0	0	1.4472	1.4472
$g_{\mathrm{Na_P}}$	0.4307	0.4307	0.4307	0.4307	0.4307	0.4307
g_{Ca}	0.54	0.54	0.54	0.54	0.54	0.54
$g_{\mathrm{K_{SS}}}$	0	0	0	0.1512	0.1512	0.1512
g_{app}	-7.1	-6.5	-7.6	-10.5	-9.8	-9.2
$g_{\mathrm{Na,SOM}}$	0	100	100	100	100	0
$g_{\mathrm{K_{DR},SOM}}$	0	80	80	80	80	0
$g_{\mathrm{leak,SOM}}$	0	0.1	0.1	0.1	0.1	0
$I_{\mathrm{app,SOM}}$	0	0.95	0.95	0.95	0.95	0
$g_{\mathrm{RS\to SOM}}$	0	0.075	0.075	0.075	0.075	0
$g_{\mathrm{SOM\to RS}}$	0	0.15	0.15	0.15	0.15	0

or in terms of its forward and backward rate functions, α_x and β_x, by the equation

$$\frac{dx}{dt} = (1-x)\alpha_x - x\beta_x.$$

Only the expressions for m_{Na} differed slightly:

$$m_{\mathrm{Na}}(V) = \alpha_m / \left(\alpha_m + \beta_m\right), \qquad m_{\mathrm{Na,SOM}}(V) = \left[1 + \exp\left((-V-38)/10\right)\right]^{-1}.$$

Steady-state values, time constants, and forward and backward rate functions are given in Table 4. For numerical stability, the backwards and forwards rate constants for q and s were converted to steady-state values and time constants before integration, using the equations

$$x_\infty = \alpha_x \tau_x, \qquad \tau_x = \left(\alpha_x + \beta_x\right)^{-1}.$$

Table 4. Activation variable dynamics.

h	$\alpha_h(V) = 0.07\exp\left(-(V+30)/20\right)$	$\beta_h(V) = \left(\exp\left(-V/10\right)+1\right)^{-1}$
m_{Na}	$\alpha_m(V) = -\frac{V+16}{10(\exp(-(V+16)/10)-1)}$	$\beta_m(V) = 4\exp\left(-(V+41)/18\right)$
$m_{\mathrm{K_{DR}}}$	$\alpha_m(V) = -0.01\frac{V+20}{\exp(-(V+20)/10)-1}$	$\beta_m(V) = 0.125\exp\left(-(V+30)/80\right)$
n	$n_\infty(V) = \left[1+\exp\left(-(V+35)/10\right)\right]^{-1}$	$\tau_n(V) = \frac{1000/(3.3*3^{(34-22)/10})}{\exp\left(\frac{V+35}{40}\right)+\exp\left(\frac{-(V+35)}{20}\right)}$
$m_{\mathrm{Na_P}}$	$m_\infty(V) = \left[1+\exp\left(-(V+40)/5\right)\right]^{-1}$	$\tau_m = 5$
s	$\alpha_s(V) = 1.6\left(1+\exp\left(-0.072(V+65)\right)\right)$	$\beta_s(V) = 0.02\frac{V+51.1}{\exp\left(\frac{V+51.1}{5}\right)-1}$
q	$\alpha_q(C_{\mathrm{Ca}}) = \min\left(0.1C_{\mathrm{Ca}},\quad 1\right)$	$\beta_q = 0.002$
h_{SOM}	$h_\infty(V) = \left[1+\exp\left((V+58.3)/6.7\right)\right]^{-1}$	$\tau_h(V) = 0.225 + 1.125\left[1+\exp\left((V+37)/15\right)\right]^{-1}$
$m_{\mathrm{K_{DR},SOM}}$	$m_\infty(V) = \left[1+\exp\left((-V-27)/11.5\right)\right]^{-1}$	$\tau_m(V) = 0.25 + 4.35\left[1+\exp\left(-\left\lvert V+10\right\rvert/10\right)\right]^{-1}.$

The dynamics of the synaptic activation variable s were given by the equation

$$\frac{ds}{dt} = -\frac{s}{\tau_{\mathrm{D}}} + \frac{1-s}{\tau_{\mathrm{R}}}\left(1 + \tanh\left(\frac{V_{\mathrm{pre}}}{10}\right)\right),$$

with time constants $\tau_{\mathrm{R}} = 0.25$ ms, $\tau_{\mathrm{D,RS\to SOM}} = 2.5$ ms, and $\tau_{\mathrm{D,SOM\to RS}} = 50$ ms. The conductance $g_{RS\to SOM}$ was selected to preserve a one-to-one spiking ratio between RS and SOM cells.

4.2 F-I curves

For these curves, we varied the level of tonic applied current I_{app} over the range from 0 to 200 Hz, in steps of 1 Hz. We measured the spiking rate for the last 5 seconds of a 6 second simulation, omitting the transient response in the first second.

4.3 Phase-locking to rhythmic, quasi-rhythmic, and speech inputs

In addition to the tonic applied current I_{app}, to measure phase-locking to rhythmic, quasi-rhythmic, and speech inputs, we introduced time-varying applied currents. These consisted of either periodic pulses (I_{PP}), variable-duration pulse trains with varied inter-pulse intervals (I_{VP}), or speech inputs (I_{speech}).

The (spike rate adjusted) phase-locking value (PLV, [124]) of the oscillator to these inputs was calculated with the expressions

$$PLV = \left(n_s|MRV|^2 - 1\right)/\left(n_s - 1\right), \qquad MRV = \frac{1}{n_s}\sum_{i=1}^{n_s}\exp\left(\sqrt{-1}\phi_I\left(t_i^s\right)\right),$$

where MRV stands for mean resultant vector, n_s is the number of spikes, t_i^s is the time of the ith spike, and $\phi_I(t)$ is the instantaneous phase of input I at frequency ω.

4.3.1 Rhythmic inputs

Periodic pulse inputs were given by the expression

$$I_{\mathrm{PP}}(t) = g_{\mathrm{PP}}\Sigma_i\chi_{\left\{|t-t_i^*|<=w(s-1)/2s\right\}}(t) * \exp\left(-(st/w)^2\right), \tag{2}$$

Table 5. Varied pulse input (I_{VP}) parameters (see *Methods: Phase-locking to rhythmic and quasi-rhythmic inputs: Inputs* for details).

Input Bandwidth ($= f_{\mathrm{high}} - f_{\mathrm{low}}$)	f_{low}	f_{high}	d_{low}	d_{high}	s_{low}	s_{high}	o_{low}	o_{high}
1	6.5	7.5	0.25	0.3	10	40	0	0.05
1.65	6.175	7.825	0.2375	0.325	10	41	0	0.1
2.3	5.85	8.15	0.225	0.35	9	41	0	0.15
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
13.35	0.325	13.675	0.0125	0.775	1	50	0	1

where $t_i^* = 2\pi\omega i$ for $i = 1, 2, ...$ is the set of times at which pulses occur, ω is the frequency, $w = 1000d/\omega$ is the pulse width given the duty cycle $d \in (0, 1)$, $*$ is the convolution operator, and s determines how square the pulse is, with $s = 1$ being roughly normal and higher s being more square. For our simulations, we took $d = 1/4$ and $s = 25$, and ω ranged over the set $\{0.25, 0.5, 1, 1.5, \ldots, 22.5, 23\}$. Input pulses were normalized so that the total (integrated) input was 1 pA/s, and were then multiplied by a conductance varying from 0 to 4 in steps of 0.1.

For I_{PP}, the instantaneous phase $\phi_I(t)$ was obtained as the angle of the complex time series resulting from the convolution of I_{PP} with a complex Morlet wavelet having the same frequency as the input and a length of 7 cycles.

4.3.2 Quasi-rhythmic inputs

Variable-duration pulse trains were given by the expression

$$I_{\mathrm{VP}}(t) = g_{\mathrm{VP}} \Sigma_i \chi_{\left\{|t - t_i^* - o_i| <= w_i \frac{(s_i - 1)}{2 s_i}\right\}}(t) * \exp\left(-\left(\frac{s_i t}{w_i}\right)^2\right), \tag{3}$$

where

$$t_i^* = \Sigma_{j=1}^{i} 1000/\omega_j,$$

the frequencies $\{\omega_i\}_1^n$ are chosen uniformly from $[f_{\mathrm{low}}, f_{\mathrm{high}}]$, the pulse width is given by $w_i = 1000d_i/\omega_i$, the duty cycles $\{d_i\}_1^n$ are chosen uniformly from $[d_{\mathrm{low}}, d_{\mathrm{high}}]$, the shape parameters $\{s_i\}_1^n$ are chosen uniformly from $[s_{\mathrm{low}}, s_{\mathrm{high}}]$, and the offsets $\{o_i\}_1^n$ are chosen uniformly from $[o_{\mathrm{low}}, o_{\mathrm{high}}]$. For our simulations, these parameters are given in Table 5.

Since I_{VP} was composed of pulses and interpulse periods of varying duration, it was not "oscillation-like" enough to employ standard wavelet and Hilbert transforms to

obtain accurate estimates of its instantaneous phase. Instead, the following procedure was used to obtain the instantaneous phase of I_{VP}. First, the times that χ_{VP} went from zero to greater than zero ($\{a_i\}_{i=1}^n$) and from greater than zero to zero ($\{b_i\}_{i=1}^n$) were obtained. Second, we specified the phase of I_{VP} on these points via the function $\phi_I^0(t)$, a piecewise constant function satisfying

$$\frac{d}{dt}\phi_I^0(t) = \sum_{i=1}^{n}\left(\frac{3\pi}{2}\delta_{a_i}(t) + \frac{\pi}{2}\delta_{b_i}(t)\right),$$

where δ is the Dirac delta function. Finally, we determined $\phi_I(t)$ from $\phi_I^0(t)$ via linear interpolation, i.e. by setting $\phi_I(t)$ to be the piecewise linear (strictly increasing) function satisfying

$$\phi_I(0) = 0, \qquad \phi_I\left(a_i\right) = \phi_I^0\left(a_i\right), \qquad \phi_I\left(b_i\right) = \phi_I^0\left(b_i\right).$$

The resulting function $\phi_I(t)$ advances by $\pi/2$ over the support of each input pulse (the support is the interval of time over which the input pulse is nonzero), and advances by $3\pi/2$ over the time interval between the supports of consecutive pulses.

4.3.3 Speech inputs

Speech inputs were comprised of 20 blindly selected sentences from the TIMIT corpus of read speech [64], which contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences. The 16 kHz speech waveform file for each sentence was processed through a model of subcortical auditory processing [65], which decomposed the input into 128 channels containing information from distinct frequency bands, reproducing the cochlear filterbank, and applied a series of nonlinear filters reflecting the computations taking place in subcortical nuclei to each channel. We selected 16 of these channels – having center frequencies of 0.1, 0.13, 0.16, 0.21, 0.26, 0.33, 0.41 0.55, 0.65, 0.82, 1.04, 1.31, 1.65, 2.07, 2.61, and 3.29 kHz – for presentation to our computational models. We varied the multiplicative gain of the resulting waveforms from 0 to 2 in steps of 0.1 to obtain inputs at a variety of strengths. Speech onset occurred after one second of simulation.

Like varied pulse inputs, speech inputs were not "oscillation-like" enough to estimate

their instantaneous phase using standard wavelet and Hilbert transforms. Thus, we used the following procedure to extract the instantaneous phase of I_{speech}. First, we calculated the power spectrum of the auditory cortical input channel derived from the speech waveform, using the Thompson multitaper method. Second, we identified peaks in the power spectrum that were at least 2 Hz apart, and used the 2nd, 3rd, and 4th largest peaks in the power spectrum to identify the frequencies of the main oscillatory modes in the θ frequency band (the largest peak in the power spectrum was in the δ frequency band for the sentences we used). Then, we bandpassed convolved the auditory input with Morlet wavelets at these three frequencies and summed the resulting complex time series, to obtain a close approximation of the θ-frequency oscillations in the input. Finally, we took the angle of this complex time series at each point in time to be the instantaneous phase of the input at that channel.

While the distribution of the (spike rate adjusted) PLV was not normal even after log transformation, the ANOVA is robust to violation of non-normality, so we compared PLV across models, sub-bands, gains, and sentences by running a 4-way ANOVA, with gain as a continuous variable. All effects were significant, and post-hoc tests for sub-bands were run to identify the optimal sub-band across models (Fig. S2). We then compared PLV values from simulations conducted with inputs from 1000 sentences at this gain and sub-band, by running a 2-way ANOVA with sentence and model as grouping variables; post-hoc model comparisons are shown in Fig. S2.

4.4 Speech Segmentation

To determine whether the activity of our models could contribute to accurate speech segmentation, we used a sum-and-threshold method to derive putative syllabic boundaries from the activity of each model. We then compared these model-derived boundaries to syllable boundaries derived from the phonemic transcriptions of each sentence, and determined how frequently model-derived boundaries occurred for each phoneme class.

4.4.1 Model-derived syllable boundaries

To determine model-derived syllable boundaries, we first divided the auditory frequency range into 8 sub-bands consisting of 16 (adjacent) channels each. For each sub-band and each model, the output from these 16 channels was used to drive the RS cells in 16 identical but unconnected versions of the model, with a multiplicative gain that varied from 0 to 2 in steps of 0.2. To approximate the effect these RS cells might have on a shared postsynaptic neuron, their time series of spiking activity, given by $\{s_i(t)\}_1^{16}$, were convolved with an exponential kernel having decay time $w_s/5$, summed over cells, and smoothed with a gaussian kernel with $\sigma = 25/4$ ms:

$$P(t) = \sum_{i=1}^{16} \left(s_i(t) * \exp\left(-\frac{5t}{w_s}\right)\right) * \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2\right).$$

The maximum of this "postsynaptic" time series during the second prior to speech input was then used to determine a threshold

$$p^* = r_{\text{thresh}} \max\{P(t)|t <= 1000\text{ms}\}$$

and the ordered set of times $\{m_i^*\}$ at which $P(t)$ crossed p^* from below were extracted as candidate syllable boundaries. Starting with $i = 2$, any candidate boundary m_i^* that followed the previous candidate boundary m_{i-1}^* with a delay less than a refractory period of 25 ms was removed from $\{m_i^*\}$ to yield a set of model-derived syllable boundaries $m = \{m_i\}_1^{n_m}$.

4.4.2 Transcription-derived syllable boundaries

Phoneme identity and boundaries have been labelled by phoneticians in every sentence of the TIMIT corpus. We used the Tsylb2 program [125] that automatically syllabifies phonetic transcriptions [126] to merge these sequences of phonemes into sequences of syllables according to English grammar rules, and thus determine the (transcription-derived) syllable boundary times $\{t_i^*\}_1^{n_s}$ for each sentence. The syllable midpoints were the set $t = \{t_i\}_1^{n_t}$ obtained by averaging successive pairs of syllable boundaries,

$$t_i = (t_i^* + t_{i+1})/2, \quad i = 1, ..., n_s - 1 =: n_t.$$

4.4.3 Comparing model- and phoneme-derived syllable boundaries

To compare the sets m and t for each sentence, we used a recursively-computed point-process metric [66]. This metric is defined by

$$d_{\mathrm{VP},\tau}(m,t) = \min_{\{m=s^1,s^2,\ldots,s^l=t\}} \Sigma_1^l C\left(s^i, s^{i+1}\right),$$

where τ is a defining timescale, and m, t, and each $s^i = \{s_1^i, \ldots, s_{n_i}^i\}$ are series of boundary times, with s^i and s^{i+1} differing by at most one boundary (which can be altered, added, or removed). The "cost" of each "move" in the chain of (series of) boundary times $s^1, s^2, \ldots, s^l$ is given by

$$C\left(s^i, s^{i+1}\right) = \begin{cases} \left|s_l^i - s_m^{i+1}\right|/\tau, & \max_{\left\{\substack{j=1,\ldots,n_i \\ k=1,\ldots,n_{i+1}}\right\}} \left|s_j^i - s_k^{i+1}\right| < \tau, \\ 1, & \text{otherwise.} \end{cases}$$

In other words, the cost of moving one boundary by a distance less than τ is less than 1, while the costs of shifting a boundary by τ or more, adding boundaries, and removing boundaries are all 1. It is helpful to note that

$$\lim_{\tau\to\infty} d_{\mathrm{VP},\tau}(m,t) = |n_m - n_t|, \qquad \lim_{\tau\to 0} d_{\mathrm{VP},\tau}(m,t) = n_m + n_t.$$

Since $d_{\mathrm{VP},\tau}(m,t)$ as defined above scales with $\max(n_m, n_t)$, we normalized this distance by the number of moves that cost less than 1, and the log-transform it, defining

$$D_{\mathrm{VP},\tau}(m,t) = \log\left(\frac{\mathrm{VP},\tau(m,t)}{\#\left\{\hat{s}^i | C\left(\hat{s}^{i-1}, \hat{s}^i\right) < 1\right\}}\right),$$

where the sequence $\{m = \hat{s}^1, \ldots, \hat{s}^l = t\}$ realizes the minimum defining $D_{\mathrm{VP},\tau}(m,t)$. Thus, $D_{\mathrm{VP},\tau}(m,t) < 0$ if each boundary in m corresponds to a distinct boundary in t shifted by less than or equal to τ, and all other things being equal, this normalized distance penalizes both missed and extra model-derived syllable boundaries. We used a timescale of $\tau = 50$ ms.

The distribution of $D_{\mathrm{VP},50}$ for each model was determined by the Kolmogorov-Smirnov test to be normal, so we compared $D_{\mathrm{VP},50}$ across models, sub-bands, gains, and sentences by running a 4-way ANOVA. All effects were significant,

and post-hoc tests for sub-bands and gains were run to identify the optimal gain and sub-band across models (Fig. S4). We then compared $D_{\mathrm{VP},50}$ values from simulations with inputs at this gain and sub-band extracted from 1000 sentences, by running a 2-way ANOVA with sentence and model as grouping variables; post-hoc tests are shown in Fig. S4.

4.4.4 Phoneme distributions of model boundaries

To determine the phoneme distributions of model boundaries, we used the phonemic transcriptions from the TIMIT corpus. The time of each model-derived boundary was compared to the set of onset and offset times of phonemes to determine the identity of the phoneme at boundary occurrence. For each simulation, we constructed a histogram over all phonemes in the TIMIT corpus; we then combined the histograms across simulations, and multiplied them by a matrix whose rows were indicator functions for 7 different phoneme classes – stops, affricates, fricatives, nasals, semivowels and glides, vowels, and other, a category which included pauses. We performed the same procedure for the set of mid-syllable times for each sentence we used in the corpus to obtain the phoneme distribution at mid-syllable.

4.5 Spike-triggered input pulses

To explore the buildup of outward current and delay of subsequent spiking induced by strong forcing, we probed each model with a single spike-triggered pulse. These pulses were triggered by the first spike after a transient interval of 2000 ms, had a pulse duration of 50 ms, and had a form given by the summand in Equation 2 with $w = 50$ and $s = 25$ (i was 1 and t_i was the time of the triggering spike).

Acknowledgments

BRPP, DAS, CES, and NJK were supported by NIH P50-MH109429. MAW was supported by Wellcome Trust #098353. CES was supported by NIH R01-MH111439. We thank Oded Ghitza and Laura Dilley for many useful discussions.

References

1. Marslen-Wilson WD. Functional parallelism in spoken word-recognition. Cognition. 1987;25(1-2):71–102.

2. Luce PA, CONOR T. 24 Spoken Word Recognition: The Challenge of Variation. The handbook of speech perception. 2005; p. 591.

3. Stevens KN. Features in speech perception and lexical access. The handbook of speech perception. 2005; p. 125–155.

4. Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. The Journal of the Acoustical Society of America. 2002;111(4):1872–1891.

5. Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech communication. 2003;41(1):245–255.

6. Ghitza O. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Frontiers in psychology. 2011;2:130.

7. Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nature neuroscience. 2012;15(4):511.

8. Ghitza O. Neuronal oscillations in decoding time-compressed speech. The Journal of the Acoustical Society of America. 2016;139(4):2190–2190.

9. Bosker HR, Ghitza O. Entrained theta oscillations guide perception of subsequent speech: behavioural evidence from rate normalisation. Language, Cognition and Neuroscience. 2018;33(8):955–967.

10. Penn LR, Ayasse ND, Wingfield A, Ghitza O. The possible role of brain rhythms in perceiving fast speech: Evidence from adult aging. The Journal of the Acoustical Society of America. 2018;144(4):2088–2094.

11. Ghitza O, Greenberg S. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica. 2009;66(1-2):113–126.

12. Ghitza O. On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Frontiers in psychology. 2012;3:238.

13. Ghitza O. Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. Frontiers in psychology. 2014;5:652.

14. Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. Trends in cognitive sciences. 2008;12(3):106–113.

15. Arnal LH, Giraud AL. Cortical oscillations and sensory predictions. Trends in cognitive sciences. 2012;16(7):390–398.

16. Ghitza O. The theta-syllable: a unit of speech information defined by cortical function. Frontiers in psychology. 2013;4:138.

17. Lewis AG, Bastiaansen M. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex. 2015;68:155–168.

18. Morillon B, Schroeder CE. Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. Annals of the New York Academy of Sciences. 2015;1337(1):26–31.

19. Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. Phil Trans R Soc Lond B. 1992;336(1278):367–373.

20. Hirst D, Di Cristo A. Intonation systems: a survey of twenty languages. Cambridge University Press; 1998.

21. Yang Lc. Duration and Pauses as Boundary-Markers in Speech: A Cross-Linguistic Study. In: Eighth Annual Conference of the International Speech Communication Association; 2007.

22. Yang X, Shen X, Li W, Yang Y. How listeners weight acoustic cues to intonational phrase boundaries. PloS one. 2014;9(7):e102166.

23. Ohala JJ. The temporal regulation of speech. Auditory analysis and perception of speech. 1975; p. 431–453.

24. Greenberg S. Speaking in shorthand–A syllable-centric perspective for understanding pronunciation variation. Speech Communication. 1999;29(2-4):159–176.

25. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA. The natural statistics of audiovisual speech. PLoS computational biology. 2009;5(7):e1000436.

26. Elliott TM, Theunissen FE. The modulation transfer function for speech intelligibility. PLoS computational biology. 2009;5(3):e1000302.

27. Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. Neuroscience & Biobehavioral Reviews. 2017;.

28. Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. The Journal of the Acoustical Society of America. 1994;95(5):2670–2680.

29. Miller GA, Licklider JC. The intelligibility of interrupted speech. The Journal of the Acoustical Society of America. 1950;22(2):167–173.

30. Huggins AWF. Distortion of the temporal pattern of speech: Interruption and alternation. The Journal of the Acoustical Society of America. 1964;36(6):1055–1064.

31. Stilp CE, Kiefte M, Alexander JM, Kluender KR. Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. The Journal of the Acoustical Society of America. 2010;128(4):2112–2126.

32. Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded

from auditory cortex. Proceedings of the National Academy of Sciences. 2001;98(23):13367–13372.

33. Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007;54(6):1001–1010.

34. Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, et al. Temporal envelope of time-compressed speech represented in the human auditory cortex. Journal of Neuroscience. 2009;29(49):15564–15574.

35. Hertrich I, Dietrich S, Trouvain J, Moos A, Ackermann H. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology. 2012;49(3):322–334.

36. Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral cortex. 2012;23(6):1378–1387.

37. Doelling KB, Arnal LH, Ghitza O, Poeppel D. Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage. 2014;85:761–768.

38. Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. Nature neuroscience. 2016;19(1):158.

39. Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E. Neural Entrainment to Speech Modulates Speech Intelligibility. Current Biology. 2017;.

40. Wilsch A, Neuling T, Herrmann CS. Envelope-tACS modulates intelligibility of speech in noise. bioRxiv. 2017; p. 097576.

41. Wilsch A, Neuling T, Obleser J, Herrmann CS. Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. NeuroImage. 2018;172:766–774.

42. Zoefel B, Archer-Boyd A, Davis MH. Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. Current Biology. 2018;.

43. Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. Journal of neurophysiology. 2005;94(3):1904–1911.

44. Shamir M, Ghitza O, Epstein S, Kopell N. Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. PLoS computational biology. 2009;5(5):e1000370.

45. Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud AL. Speech encoding by coupled cortical theta and gamma oscillations. Elife. 2015;4.

46. Räsänen O, Doyle G, Frank MC. Pre-linguistic segmentation of speech into syllable-like units. Cognition. 2018;171:130–150.

47. Hovsepyan S, Olasagasti I, Giraud AL. Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. Nature communications. 2020;11(1):1–12.

48. Rotstein HG, Pervouchine DD, Acker CD, Gillies MJ, White JA, Buhl EH, et al. Slow and fast inhibition and an H-current interact to create a theta rhythm in a model of CA1 interneuron network. Journal of neurophysiology. 2005;94(2):1509–1518.

49. Gutfreund Y, Segev I, et al. Subthreshold oscillations and resonant frequency in guinea-pig cortical neurons: physiology and modelling. The Journal of physiology. 1995;483(3):621–640.

50. Carracedo LM, Kjeldsen H, Cunnington L, Jenkins A, Schofield I, Cunningham MO, et al. A neocortical delta rhythm facilitates reciprocal interlaminar interactions via nested theta rhythms. Journal of Neuroscience. 2013;33(26):10750–10761.

51. Cannon J, Kopell N. The leaky oscillator: Properties of inhibition-based rhythms revealed through the singular phase response curve. SIAM Journal on Applied Dynamical Systems. 2015;14(4):1930–1977.

52. Sherfey JS, Ardid S, Hass J, Hasselmo ME, Kopell NJ. Flexible resonance in prefrontal networks with strong feedback inhibition. PLoS computational biology. 2018;14(8):e1006357.

53. Ghitza O. "Acoustic-driven oscillators as cortical pacemaker": a commentary on Meyer, Sun & Martin (2019). Language, Cognition and Neuroscience. 2020; p. 1–6.

54. Ermentrout GB. n: m Phase-locking of weakly coupled oscillators. Journal of Mathematical Biology. 1981;12(3):327–342.

55. Ermentrout B. Type I membranes, phase resetting curves, and synchrony. Neural computation. 1996;8(5):979–1001.

56. Kopell N, Ermentrout G. Mechanisms of phase-locking and frequency control in pairs of coupled neural oscillators. Handbook of dynamical systems. 2002;2:3–54.

57. Achuthan S, Canavier CC. Phase-resetting curves determine synchronization, phase locking, and clustering in networks of neural oscillators. Journal of Neuroscience. 2009;29(16):5218–5233.

58. Canavier CC, Achuthan S. Pulse coupled oscillators and the phase resetting curve. Mathematical biosciences. 2010;226(2):77–96.

59. Klinshov V, Yanchuk S, Stephan A, Nekorkin V. Phase response function for oscillators with strong forcing or coupling. EPL (Europhysics Letters). 2017;118(5):50006.

60. Canavier CC, Kazanci FG, Prinz AA. Phase resetting curves allow for simple and accurate prediction of robust N: 1 phase locking for strongly coupled neural oscillators. Biophysical journal. 2009;97(1):59–73.

61. Zhou Y, Vo T, Rotstein HG, McCarthy MM, Kopell N. M-Current Expands the Range of Gamma Frequency Inputs to Which a Neuronal Target Entrains. The Journal of Mathematical Neuroscience. 2018;8(1):13.

62. Ten Oever S, Sack AT. Oscillatory phase shapes syllable perception. Proceedings of the National Academy of Sciences. 2015;112(52):15833–15837.

63. Adams NE, Teige C, Mollo G, Karapanagiotidis T, Cornelissen PL, Smallwood J, et al. Theta/delta coupling across cortical laminae contributes to semantic cognition. Journal of neurophysiology. 2019;121(4):1150–1161.

64. Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS. DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1. DARPA; 1993.

65. Chi T, Ru P, Shamma SA. Multiresolution spectrotemporal analysis of complex sounds. The Journal of the Acoustical Society of America. 2005;118(2):887–906.

66. Victor JD, Purpura KP. Metric-space analysis of spike trains: theory, algorithms and application. Network: computation in neural systems. 1997;8(2):127–164.

67. Ermentrout B, Pascal M, Gutkin B. The effects of spike frequency adaptation and negative feedback on the synchronization of neural oscillators. Neural computation. 2001;13(6):1285–1310.

68. Acker CD, Kopell N, White JA. Synchronization of strongly coupled excitatory neurons: relating network behavior to biophysics. Journal of computational neuroscience. 2003;15(1):71–90.

69. Hu H, Vervaeke K, Storm JF. Two forms of electrical resonance at theta frequencies, generated by M-current, h-current and persistent Na+ current in rat hippocampal pyramidal cells. The Journal of physiology. 2002;545(3):783–805.

70. Rotstein HG, Nadim F. Frequency preference in two-dimensional neural models: a linear analysis of the interaction between resonant and amplifying currents. Journal of computational neuroscience. 2014;37(1):9–28.

71. Rotstein HG. Spiking resonances in models with the same slow resonant and fast amplifying currents but different subthreshold dynamic properties. Journal of computational neuroscience. 2017;43(3):243–271.

72. Akam TE, Kullmann DM. Efficient "communication through coherence" requires oscillations structured to minimize interference between signals. PLoS computational biology. 2012;8(11):e1002760.

73. Tsai TYC, Choi YS, Ma W, Pomerening JR, Tang C, Ferrell JE. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. Science. 2008;321(5885):126–129.

74. Atallah BV, Scanziani M. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. Neuron. 2009;62(4):566–577.

75. Shin D, Cho KH. Recurrent connections form a phase-locking neuronal tuner for frequency-dependent selective communication. Scientific reports. 2013;3:2519.

76. Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE. The spectrotemporal filter mechanism of auditory selective attention. Neuron. 2013;77(4):750–761.

77. Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S. Rhythmic auditory cortex activity at multiple timescales shapes stimulus–response gain and background firing. Journal of Neuroscience. 2015;35(20):7750–7762.

78. Teng X, Tian X, Doelling K, Poeppel D. Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. European Journal of Neuroscience. 2017;.

79. Ghitza O. Acoustic-driven delta rhythms as prosodic markers. Language, Cognition and Neuroscience. 2017;32(5):545–561.

80. Stanley DA, Falchier AY, Pittman-Polletta BR, Lakatos P, Whittington MA, Schroeder CE, et al. Flexible reset and entrainment of delta oscillations in primate primary auditory cortex: modeling and experiment. bioRxiv. 2019; p. 812024.

81. Ahissar E, Ahissar M. 18. Processing of the temporal envelope of speech. The auditory cortex: A synthesis of human and animal research. 2005; p. 295.

82. Dilley LC, Pitt MA. Altering context speech rate can cause words to appear or disappear. Psychological Science. 2010;21(11):1664–1670.

83. Dilley LC, Mattys SL, Vinke L. Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. Journal of Memory and Language. 2010;63(3):274–294.

84. Brown M, Salverda AP, Dilley LC, Tanenhaus MK. Expectations from preceding prosody influence segmentation in online sentence processing. Psychonomic bulletin & review. 2011;18(6):1189–1196.

85. Baese-Berk MM, Heffner CC, Dilley LC, Pitt MA, Morrill TH, McAuley JD. Long-term temporal tracking of speech rate affects spoken-word recognition. Psychological Science. 2014;25(8):1546–1553.

86. Brown M, Salverda AP, Dilley LC, Tanenhaus MK. Metrical expectations from preceding prosody influence perception of lexical stress. Journal of Experimental Psychology: Human Perception and Performance. 2015;41(2):306.

87. Kösem A, Bosker HR, Takashima A, Meyer AS, Jensen O, Hagoort P. Neural entrainment determines the words we hear. 2017;.

88. Christiansen MH, Chater N. The Now-or-Never bottleneck: A fundamental constraint on language. Behavioral and Brain Sciences. 2016;39.

89. Pefkou M, Arnal LH, Fontolan L, Giraud AL. θ-Band and β-Band Neural Activity Reflects Independent Syllable Tracking and Comprehension of Time-Compressed Speech. Journal of Neuroscience. 2017;37(33):7930–7938.

90. Riecke L, Sack AT, Schroeder CE. Endogenous delta/theta sound-brain phase entrainment accelerates the buildup of auditory streaming. Current Biology. 2015;25(24):3196–3201.

91. Riecke L, Formisano E, Herrmann CS, Sack AT. 4-Hz transcranial alternating current stimulation phase modulates hearing. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation. 2015;8(4):777–783.

92. Hamilton LS, Edwards E, Chang EF. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. Current Biology. 2018;28(12):1860–1871.

93. Oganian Y, Chang EF. A speech envelope landmark for syllable encoding in human superior temporal gyrus. Science advances. 2019;5(11):eaay6279.

94. O'connell M, Barczak A, Ross D, McGinnis T, Schroeder C, Lakatos P. Multi-scale entrainment of coupled neuronal oscillations in primary auditory cortex. Frontiers in human neuroscience. 2015;9:655.

95. Henry MJ, Obleser J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proceedings of the National Academy of Sciences. 2012;109(49):20095–20100.

96. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS biology. 2013;11(12):e1001752.

97. Horton C, D'Zmura M, Srinivasan R. Suppression of competing speech through entrainment of cortical oscillations. Journal of neurophysiology. 2013;109(12):3082–3093.

98. Ding N, Simon JZ. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. Journal of Neuroscience. 2013;33(13):5728–5735.

99. Yellamsetty A, Bidelman GM. Low-and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. Hearing research. 2018;361:92–102.

100. Oribe N, Onitsuka T, Hirano S, Hirano Y, Maekawa T, Obayashi C, et al. Differentiation between bipolar disorder and schizophrenia revealed by neural oscillation to speech sounds: an MEG study. Bipolar disorders. 2010;12(8):804–812.

101. Soltész F, Szűcs D, Leong V, White S, Goswami U. Differential entrainment of neuroelectric delta oscillations in developmental dyslexia. PLoS One. 2013;8(10):e76608.

102. Jochaut D, Lehongre K, Saitovitch A, Devauchelle AD, Olasagasti I, Chabane N, et al. Atypical coordination of cortical oscillations in response to speech in autism. Frontiers in human neuroscience. 2015;9:171.

103. Wieland EA, McAuley JD, Dilley LC, Chang SE. Evidence for a rhythm perception deficit in children who stutter. Brain and language. 2015;144:26–34.

104. Jiménez-Bravo M, Marrero V, Benítez-Burraco A. An oscillopathic approach to developmental dyslexia: From genes to speech processing. Behavioural brain research. 2017;329:84–95.

105. Di Liberto GM, O'Sullivan JA, Lalor EC. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Current Biology. 2015;25(19):2457–2465.

106. Mai G, Minett JW, Wang WSY. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. Neuroimage. 2016;133:516–528.

107. Ding N, Chatterjee M, Simon JZ. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage. 2014;88:41–46.

108. Zoefel B, VanRullen R. The role of high-level processes for oscillatory phase entrainment to speech sound. Frontiers in human neuroscience. 2015;9:651.

109. Zoefel B, VanRullen R. EEG oscillations entrain their phase to high-level features of speech sound. Neuroimage. 2016;124:16–23.

110. Park H, Ince RA, Schyns PG, Thut G, Gross J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Current Biology. 2015;25(12):1649–1653.

111. Keitel A, Ince RA, Gross J, Kayser C. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. NeuroImage. 2017;147:32–42.

112. Keitel A, Gross J, Kayser C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS biology. 2018;16(3):e2004473.

113. Hasselmo ME, McGaughy J. High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. Progress in brain research. 2004;145:207–231.

114. Hasselmo ME. The role of acetylcholine in learning and memory. Current opinion in neurobiology. 2006;16(6):710–715.

115. Honey CJ, Newman EL, Schapiro AC. Switching between internal and external modes: a multiscale learning principle. Network Neuroscience. 2017;1(4):339–356.

116. McFarland WL, Teitelbaum H, Hedges EK. Relationship between hippocampal theta activity and running speed in the rat. Journal of comparative and physiological psychology. 1975;88(1):324.

117. Kleinfeld D, Ahissar E, Diamond ME. Active sensation: insights from the rodent vibrissa sensorimotor system. Current opinion in neurobiology. 2006;16(4):435–444.

118. Kleinfeld D, Deschenes M, Ulanovsky N. Whisking, sniffing, and the hippocampal θ-rhythm: a tale of two oscillators. PLoS biology. 2016;14(2):e1002385.

119. Groh A, Meyer HS, Schmidt EF, Heintz N, Sakmann B, Krieger P. Cell-type specific properties of pyramidal neurons in neocortex underlying a layout that is modifiable depending on the cortical area. Cerebral cortex. 2010;20(4):826–836.

120. Kim EJ, Juavinett AL, Kyubwa EM, Jacobs MW, Callaway EM. Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. Neuron. 2015;88(6):1253–1267.

121. Sherfey JS, Soplata AE, Ardid S, Roberts EA, Stanley DA, Pittman-Polletta BR, et al. DynaSim: a MATLAB Toolbox for neural modeling and simulation. Frontiers in neuroinformatics. 2018;12:10.

122. Traub RD, Wong RK, Miles R, Michelson H. A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. Journal of Neurophysiology. 1991;66(2):635–650.

123. Lee JH, Whittington MA, Kopell NJ. Top-down beta rhythms support selective attention via interlaminar interaction: a model. PLoS computational biology. 2013;9(8):e1003164.

124. Aydore S, Pantazis D, Leahy RM. A note on the phase locking value and its properties. NeuroImage. 2013;74:231–244. doi:10.1016/j.neuroimage.2013.02.008.

125. Fisher W. Program TSYLB (version 2 revision 1.1); 1996.

126. Kahn D. Syllable-based generalizations in English. Bloomington: Indiana. 1976;.

Supplemental Material

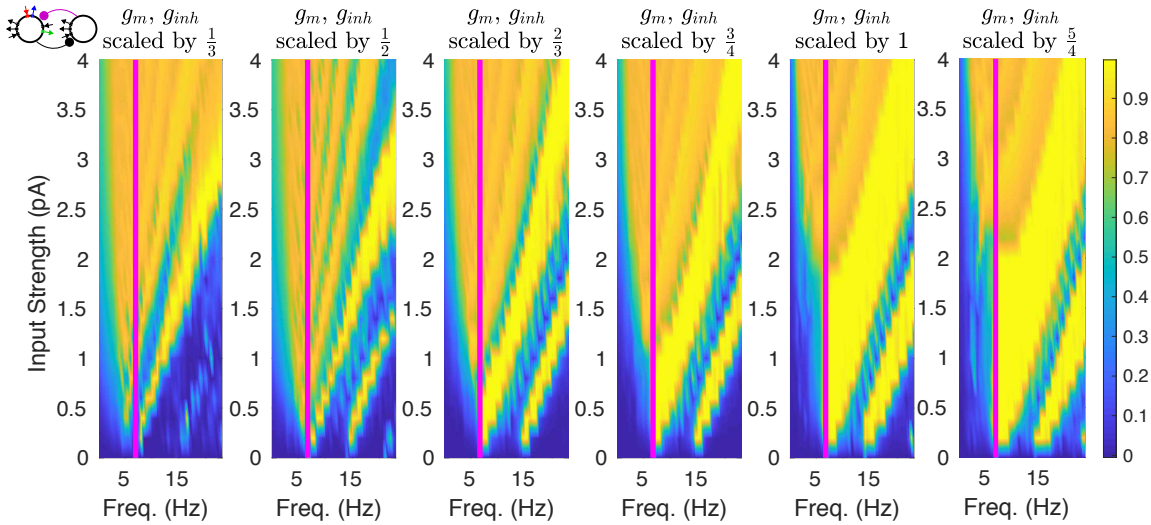

Fig S1. Dependence of One-to-One Phase Locking on Inhibitory Conductance. We multiplied the conductances g_{m} and g_{inh} in model MIS by factors of $\frac{1}{3}$, $\frac{1}{2}$, $\frac{3}{4}$, 1, and $\frac{5}{4}$, and then computed plots of PLV for different input frequencies and strengths, as in Fig. 3. The bright yellow band in each figure, representing the region of one-to-one phase-locking, depends on the size of g_{m} and g_{inh}; both increase from left to right.

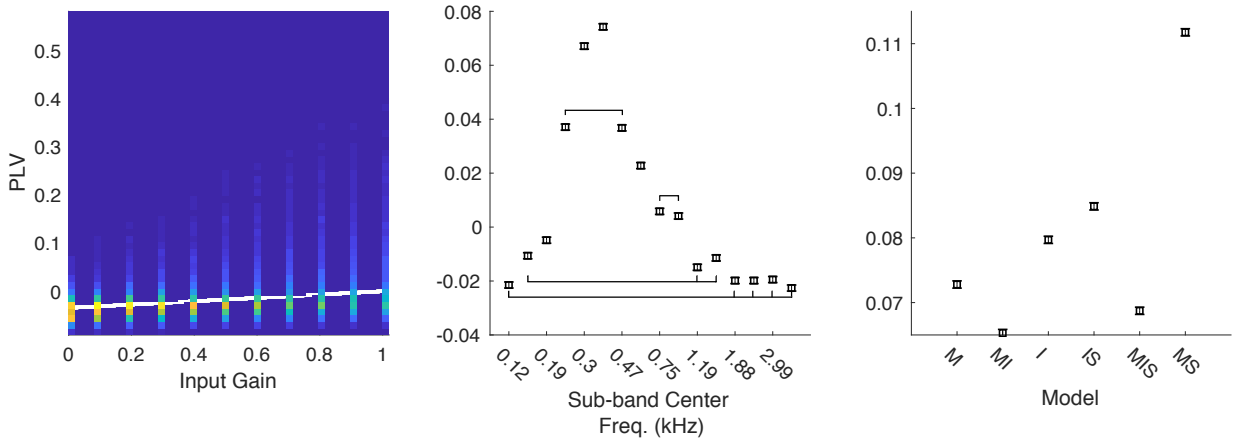

Fig S2. Statistical Tests of PLV. PLV depended linearly on input gain (left), as shown by a plot of the joint density of input gain and PLV, along with the regression line of PLV onto input gain (white, $p < 10^{-10}$). In an ANOVA with gain treated as a continuous regressor, the group effect for channels was highly significant (middle, $p < 10^{-10}$); lines connect channels that are *not* significantly different in post-hoc tests at level $\alpha = .05$. In a separate ANOVA for results from simulations with input from 1000 sentences at only the optimal gain and channel, post-hoc tests showed significant differences between all models at level $\alpha = .05$.

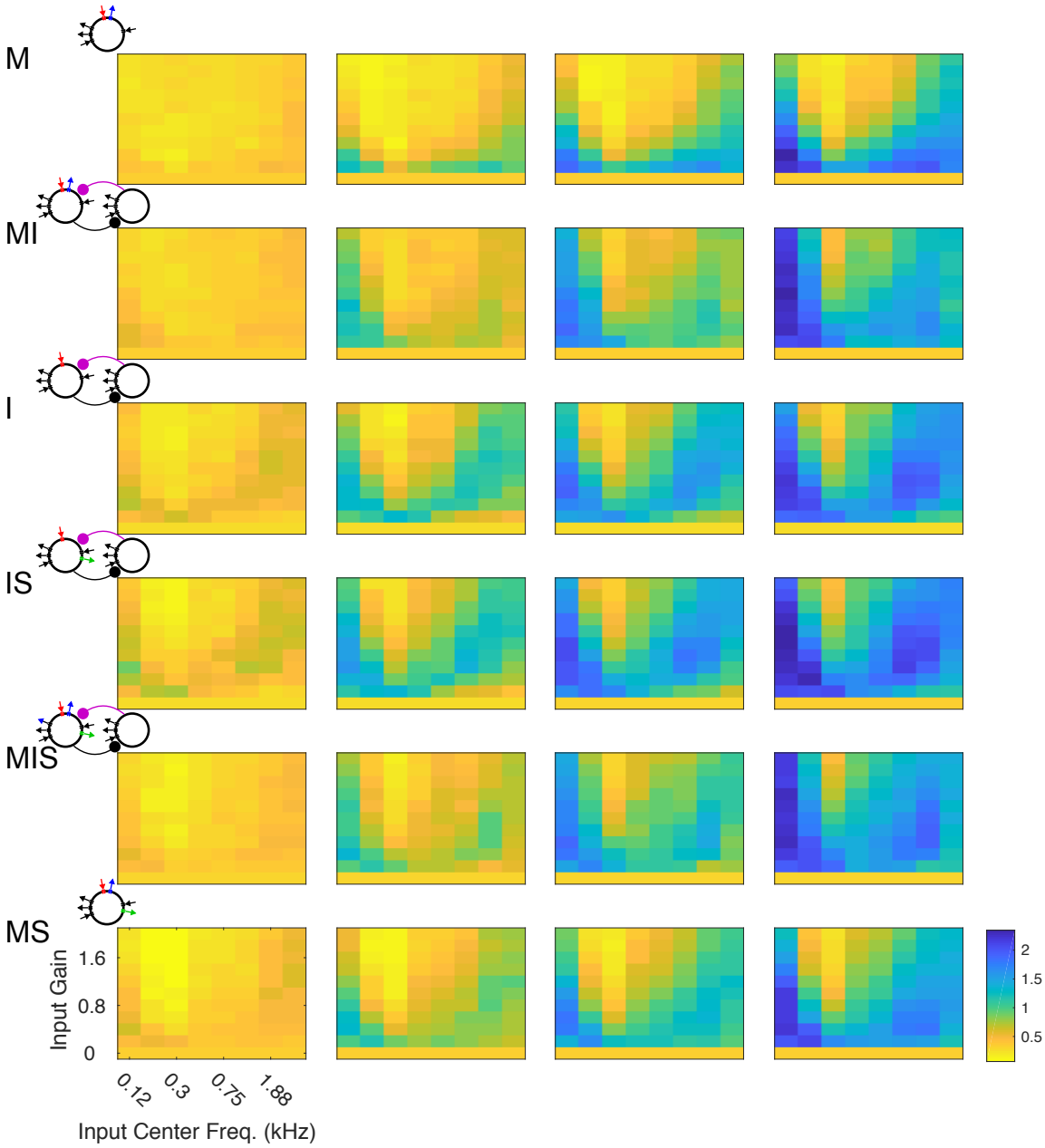

Fig S3. Segmentation Performance Depends on Threshold. False-color plots show the mean $D_{\mathrm{VP},50}$ for different auditory sub-bands (x-axis) as well as varying input strengths (y-axis) for all six models, with model-derived boundaries determined by the parameters $w_s = 75$ and $r_{\mathrm{thresh}} = 1/3$ (left), $r_{\mathrm{thresh}} = 0.45$ (middle left), $r_{\mathrm{thresh}} = 0.55$ (middle right), and $r_{\mathrm{thresh}} = 2/3$. The model exhibiting the best segmentation performance shifts with the value of r_{thresh}.

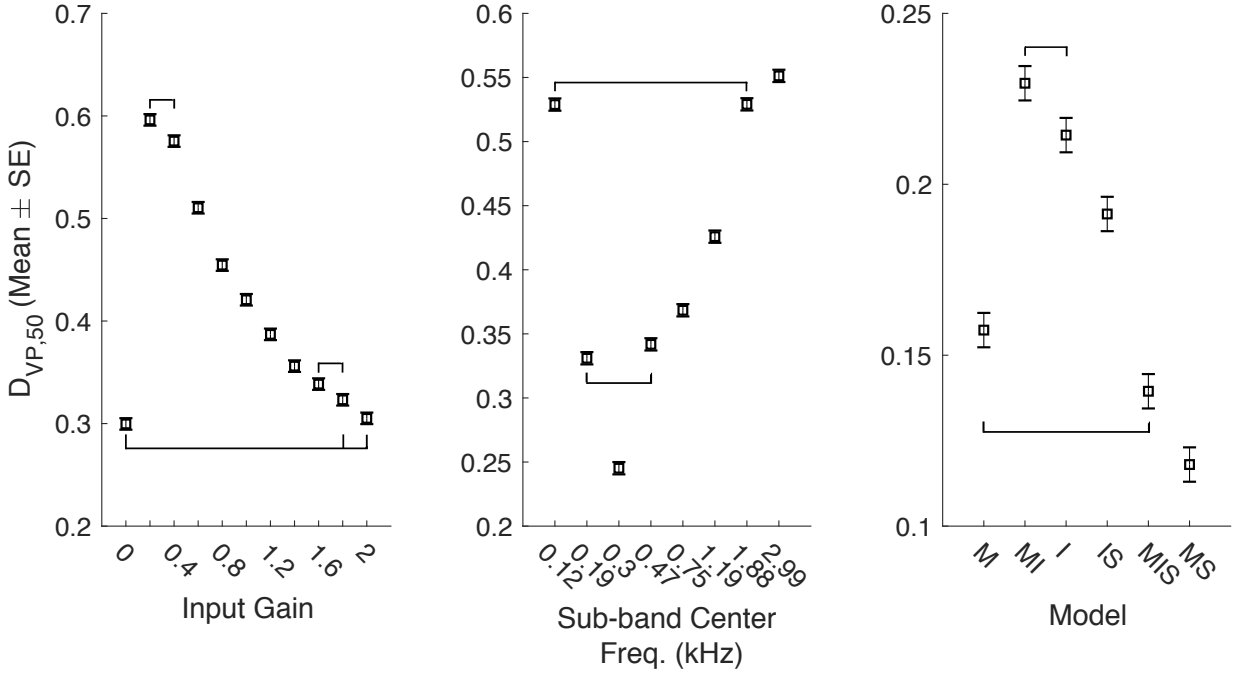

Fig S4. Statistical Tests of $D_{\mathbf{VP},50}$. In an ANOVA treating input gain (left), sub-band center frequency (middle), and model as categorical variables, all effects were highly significant ($p < 10^{-10}$). Lines connect channels that are *not* significantly different in post-hoc tests at level $\alpha = .05$. In a separate ANOVA for results from simulations with input from 1000 sentences at only the optimal gain and channel, post-hoc tests clustered the models in four groups at level $\alpha = .05$ (right).

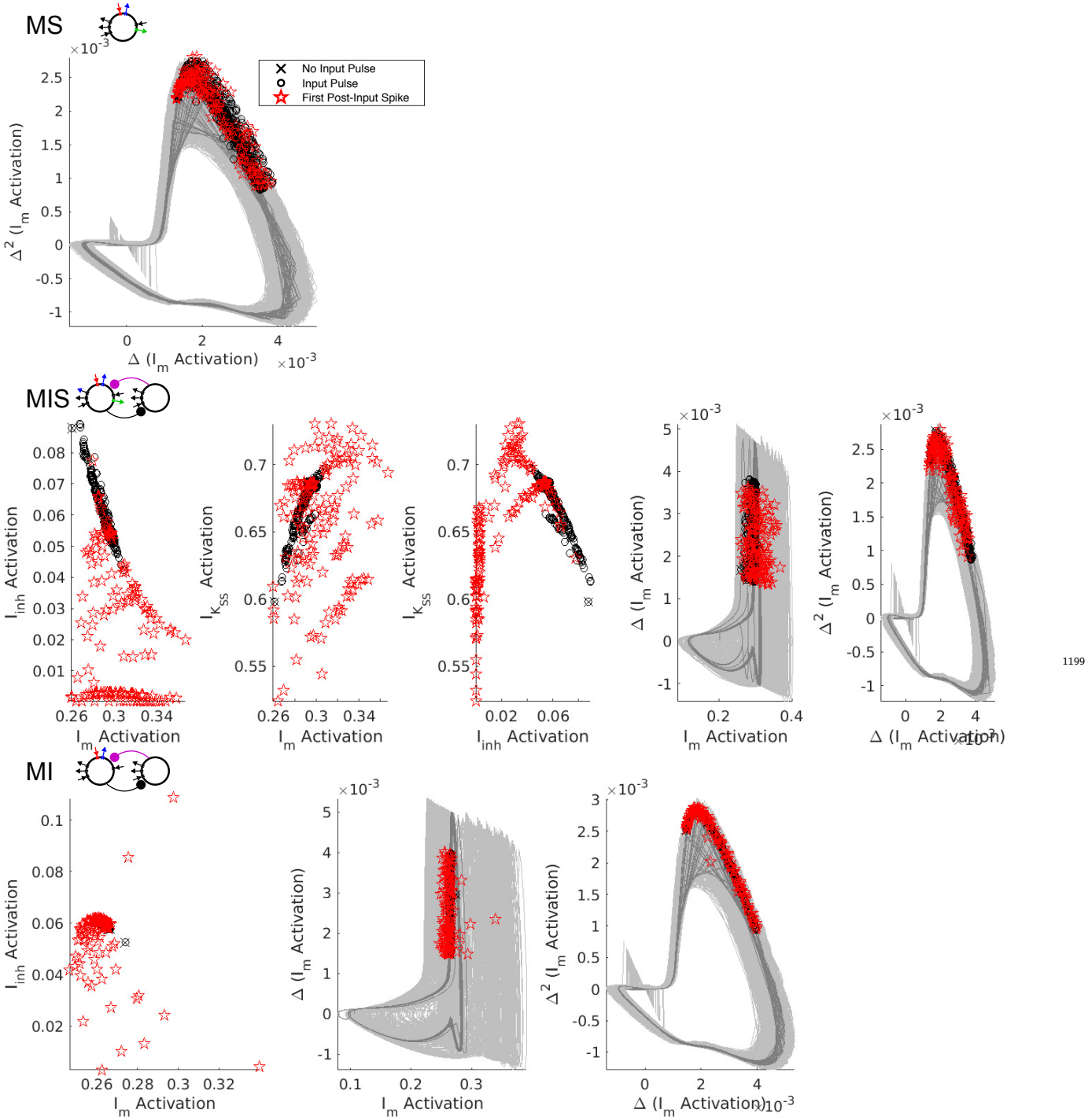

Fig S5. Dynamics of Inhibitory Currents in Models MIS and MI. Plots of the pre-spike gating variables in models MS, MIS, and MI. Top row, plotting the second difference in m-current activation level of against its first difference reveals that pre-spike activation levels are clustered along a single branch of the oscillator's trajectory. Middle row, plots of the relationships between the pre-spike activation levels of I_{inh}, I_{m}, and $I_{\mathrm{K_{SS}}}$ in model MIS, revealing a dependence on the phase of oscillations in m-current activation. Bottom, plots of the relationships between the pre-spike activation levels of I_{inh} and I_{m} in model MI, again revealing a dependence on the phase of oscillations in m-current activation. (For all plots, light gray curves represent trajectories with an input pulse; dark gray curves represent trajectories without an input pulse.)

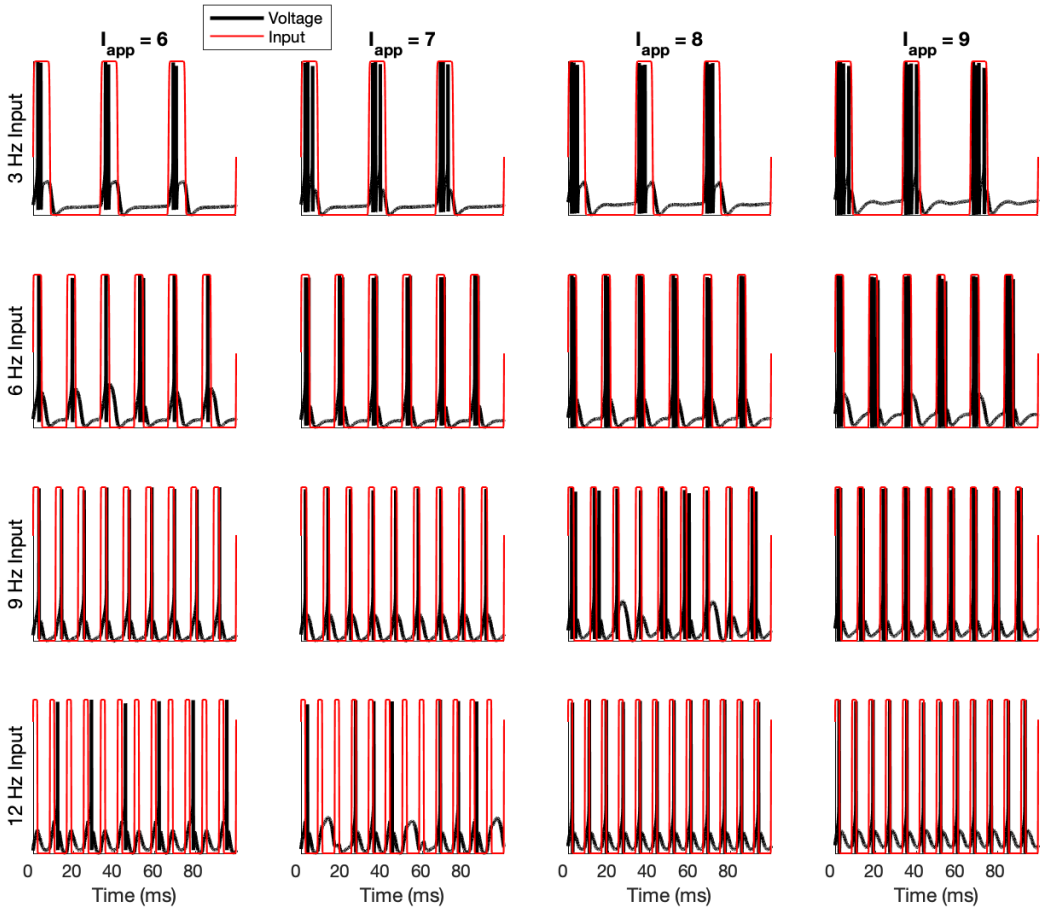

Fig S6. Varying Tonic Input to Model MS. We altered the tonic input strength g_{app} to model MS, and gave periodic pulse inputs of strength $g_{\text{PP}} = 1$ at varying frequencies. For lower levels of tonic input, phase-locking is closer to one-to-one for low frequency inputs, but many high frequency input cycles are "missed"; for higher levels of tonic input, phase-locking is one-to-one for high frequency inputs, but many-to-one for low frequency inputs.