

# USING PERIPHERAL MODELS TO IMPROVE RESPONSES TO SPEECH IN COMPUTATIONAL MODELS OF THE AUDITORY CORTEX

NICHOLAS CIMASZEWSKI  
BENJAMIN RICHARDSON  
*Department of Linguistics, University of Rochester*

## 1 Abstract

Computational models of the auditory cortex aim to replicate the brain's responses to many stimuli, and one of the hardest goals is encoding the rich and varied structure of speech in a robust way. However, often these models oversimplify the modeling of the auditory periphery, approximating them with linear filterbanks. In an attempt to improve the coding of speech stimuli, we adapt a recent two stream model for auditory cortex by augmenting it with a physiologically realistic model of the inner ear, brainstem and inner ear that accurately models the distinctive nonlinearities of these mechanisms ignored by linear approximations. We found that adding this more complex periphery results in spectrally richer responses to vowels and connected speech and improves the model's ability to encode key features of speech such as higher formant frequencies. These results suggest that future modeling work of the auditory system should emphasize the physiological realism of all stages to best study the coding of complex stimuli such as speech.

## 2 Introduction

Encoding speech is one of the most important functions of the human auditory system. Phonetic analysis has traditionally characterized speech sounds by features of their frequency content, but the ability to perceive speech is robust to dramatic changes of sound level and pitch (Hillenbrand et al. 1995, Ladefoged 2012). Thus while the analysis of speech waveforms typically characterizes vowels in terms of their formant frequencies, the brain obviously computes a representation of

speech sounds that is robust to variations across speaker and sonic conditions and not linearly dependent on formants.

Several computational models of the auditory nerve, the inferior colliculus, and auditory cortex have been developed in order to study the spike rate responses of neurons in the human auditory system and their patterns (Zilany, Bruce, Nelson, and Carney 2009, Zulfiqar et al. 2020). With models as these, it is possible to study physiologically accurate simulated responses to speech sounds. In this work, a computational model for four areas of the auditory cortex is modified to include peripheral processing models for the auditory nerve and the inferior colliculus, and the response properties to speech sounds with and without peripheral processing are studied. It is hypothesized that introducing peripheral models that are more physiologically accurate will improve cortical model responses to speech sounds, specifically vowels and a set of sentences. Modified models will allow insight into not only the models themselves, but will provide a tool for studying speech sounds at a more fundamental level than that of acoustic analysis.

## 3 Background

### 3.1 Speech Processing in the Auditory Periphery

Figure 1 shows a simple diagram of the human auditory pathway leading up to the auditory cortex. Pressure variations in the air are transmitted to the cochlea via the middle ear, and the physical vibration of the basilar membrane are transduced to electrical signals via the inner and outer hair cells (I/OHC), at which point key non-linearities are introduced. These cochlear hair cells are tuned to different frequencies, creating a tonotopic organization that is maintained throughout the auditory pathway. Information then travels from the cochlea through the auditory nerve (AN), shown in figure 1 as the three lines extending from the cochlea. The auditory nerve (AN) transmits this signal to the cochlear nucleus, and the information passes through other areas of the brainstem before reaching the inferior colliculus (IC). All auditory information must then pass through the inferior colliculus (IC), essentially a midbrain bottleneck of information before the signal is passed onto higher processing centers in the auditory cortex (AC), where the sound is understood in context. This work focuses primarily on three components of the pathway: the auditory nerve (AN), the inferior colliculus (IC), and auditory cortex (AC).

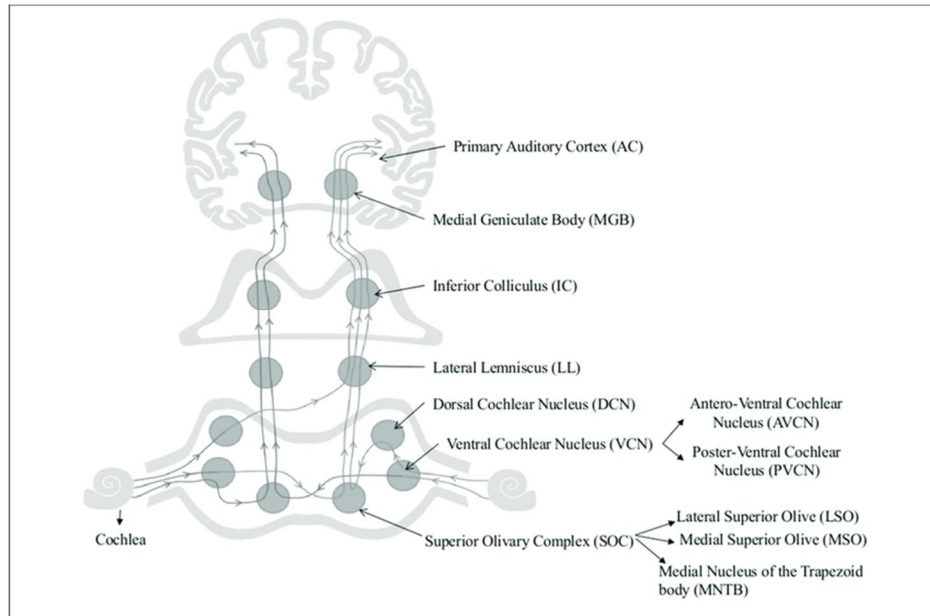


Figure 1. Simplified diagram of the human auditory pathway (from Dona 2018)

### 3.2 Computational Models

Each level of processing in the human auditory pathway encodes different features of a sound, and these patterns are fairly well defined for speech sounds. In response to harmonic sounds, AN-fiber responses are phase-locked to the beating between harmonics that creates a strong periodicity at  $F_0$ , which is the pitch of the speaker. AN fibers which are tuned to a frequency near a formant are dominated by the harmonic that is closest to the formant peak (Delgutte and Kiang 1984). The pitch of a speaker is encoded similarly in the auditory nerve (Zilany et al., 2009). Neurons in the inferior colliculus exhibit a selectivity to sound fluctuations in addition to energy at a given frequency (Carney and McDonough 2019). There are three types of tuning represented in the IC: band-enhanced (BE) cells, which are excited by amplitude modulation of attuned frequency, band-suppressed (BS) cells which are similarly inhibited, and hybrids of BE and BS. At vowel formant frequencies, BE cells show lower spike rates, and BS cells show increased spike rates (Carney and McDonough 2019).

Figure 2 shows a system diagram underlying the AN model. The stimulus is first processed through a filter that matches frequency shaping in the middle ear, and then through models of the IHC and OHC. Finally, the synapse between the inner hair cells and the auditory nerve is modeled by a power-law system, matched to patterns of neurotransmitter transmission into the synaptic cleft (Zilany, Bruce, Nelson, and Carney 2009). Figure 3 shows a schematic diagram of the Same-Frequency Inhibition-Excitation (SFIE) model for the IC for a single AN fiber (Nelson and Carney 2004).

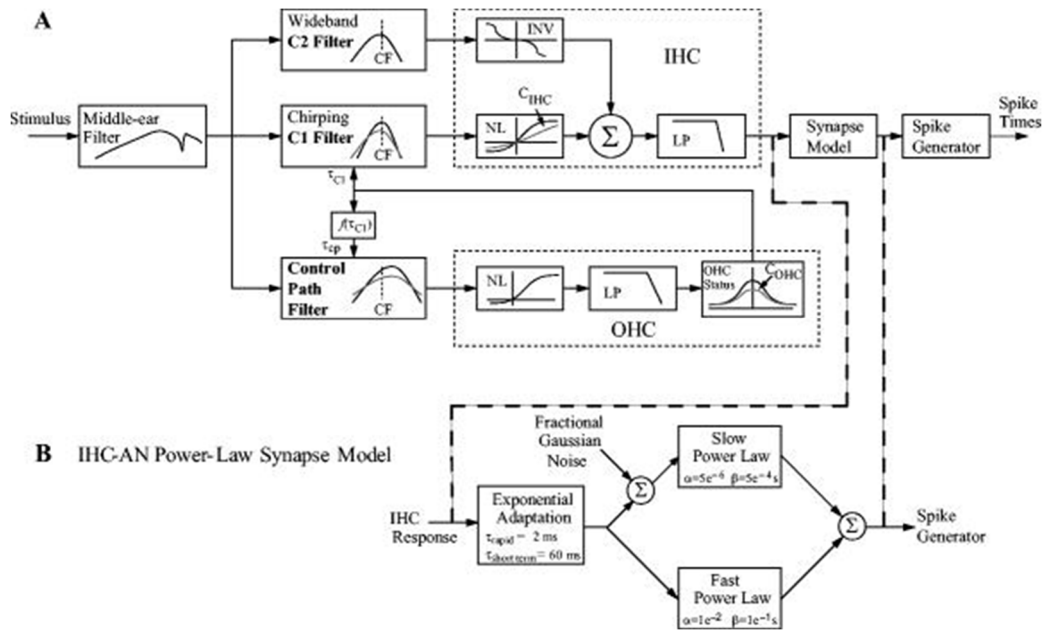


Figure 2. Schematic for auditory nerve model. From Zilany, Bruce, Nelson, and Carney 2009

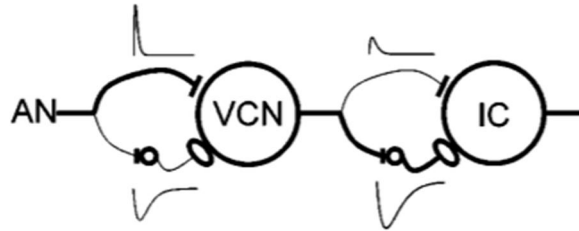


FIG. 2. Schematic diagram of the same-frequency inhibition and excitation (SFIE) model. A single model AN fiber provides the postsynaptic cell with both excitatory and inhibitory input, via an inhibitory interneuron. The thickness of the lines corresponds to the relative strength of the inhibition and excitation at each level. Alpha functions representing the assumed membrane and synaptic properties are also shown.

Figure 3. Schematic diagram of the SFIE model. From Nelson and Carney 2004, Figure 2.

Inputting a sound waveform through both of the models explained above gives us helpful information about the representation of speech sounds in the brain. Figure 4 below shows example responses in the AN and IC band-enhanced models for the center 150 ms of an /æ/ vowel spoken by a female speaker from Boston, Massachusetts. Very distinct features in these responses arise at the pitch of the speaker ( $F_0$ ) below 500 Hz and at each of the first two formant frequencies. The strong alternating bands of light and dark far below 500 Hz (around 100 or 200 Hz) in both responses represent the fact that the firing rate of neurons in these areas matches fluctuations at the pitch of the speaker, known as phase locking (Carney, Li and McDonough 2015). Furthermore, at the first and second formants, there are bands of increased activity in the AN model, and decreased activity in the IC band-enhanced model. At these frequencies, neurons phase lock to fluctuations in the sound waveform as they do to the pitch, though these bands are closer together because the component is at a higher frequency. In the auditory nerve, this is encoded by an increase in firing rates in neurons near formant frequencies. In band-enhanced cells, this is

represented by a decrease in activity as shown by the dark bands near formant frequencies. These models very strongly represent important features of vowels, and consequently other speech sounds. Therefore, they prove useful in distinguishing differences in speech sounds. When used in conjunction with cortical models, then, we would expect that we can use the output of cortex to differentiate different vowels and analyze important features.

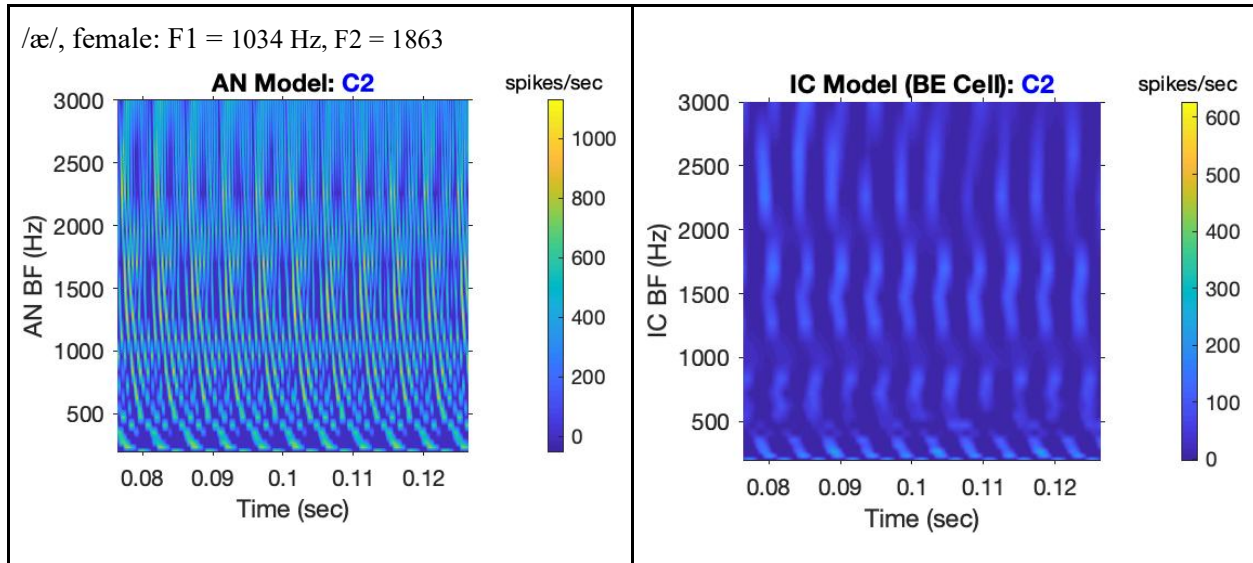


Figure 4. Example responses in the AN model and IC model for a band-enhanced cell to the vowel /æ/ for a female speaker from Boston, Massachusetts.

The main aim of this work is to use the above models for the AN and the IC to improve responses to speech in a model of auditory cortex developed by Zulfiqar and colleagues that emphasizes dorsal and ventral streams, with the goal of differentiating information encoded in a “Fast” area and a “Slow” area (Zulfiqar et al. 2020). Figure 5 shows the anatomical and systems structure of the cortical model. This model represents the activity of cortex in 4 “areas”, including two “core” areas – primary auditory cortex (A1) and a rostral area (R) – which feed into the “Fast” belt area and the “Slow” belt area respectively. Each area is composed of 98 units, which do not necessarily represent individual neurons as much as populations of neurons with similar properties.

The division of auditory processing into two streams is thought to play a significant role in auditory cognition and has been linked to deriving semantic information (“what” processing) or processing sound location and sound movement (“where” processing) (Kaas et al. 1999, Romanski et al. 1999, Belin and Zatorre 2000, Kaas and Hackett 2000, Rauschecker and Tian 2000, Tian et al. 2001, Arnott et al. 2004). The cortical areas in this model respond to both higher energy and higher modulation in lower processing areas. The information from the A1 core area feeds into the Fast belt area, and the information from the R core area feeds into the Slow belt area. The Fast and Slow areas differ in their latency of response.

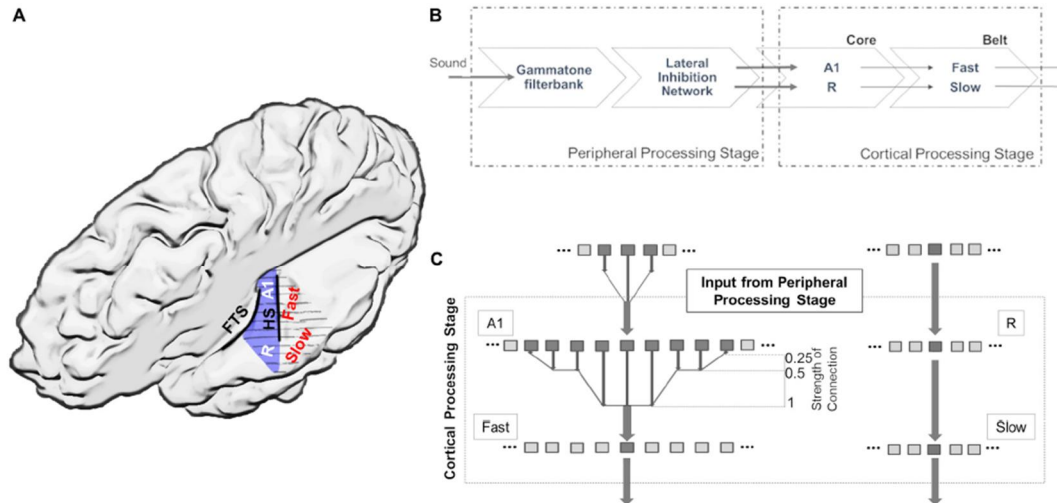


Figure 5. Structure of the model of auditory cortex from Zulfiqar et al. 2020.

Each unit in the 4 areas were simulated via a Wilson and Cowan cortical model (WCCM), which provides a physiologically reasonable firing rate output based on excitation and inhibition dynamics in the human neocortex (Wilson and Cowan 1972, 1973). The two-stream model outputs rate vs. time data for each of the 98 units in response to an auditory stimulus. Each area of the cortical model has a population of excitatory and inhibitory units, with recurrent feedback connections between all units and a common external stimulus for every unit each population. For their experiments, Zulfiqar et al. (2020) presented an inhibitory input of zero, so for the purposes of fair comparison we followed suit.

As shown in Figure 5, the peripheral processing stage in the original cortical model involves a simple bandpass filter bank and a lateral inhibition network, meant to represent broadly processing in the cochlea and subsequent brainstem and midbrain areas. We posit that improving this peripheral processing stage with physiologically reasonable models of brainstem and midbrain areas will improve the cortical responses to speech, and lead to valuable insights in speech perception and auditory cognition.

## 4 Materials and Methods

Shown below in figure 6 is a schematic of the experimental design. In order to compare how the addition of different peripheral models changes responses to speech sounds, recorded wav files for a male and female speaker were run through three different versions of the cortical model, as follows: 1) The cortical model with the original simple periphery, 2) the cortical model receiving input directly from the auditory nerve response, and 3) the cortical model receiving input directly from the IC BE cell response.

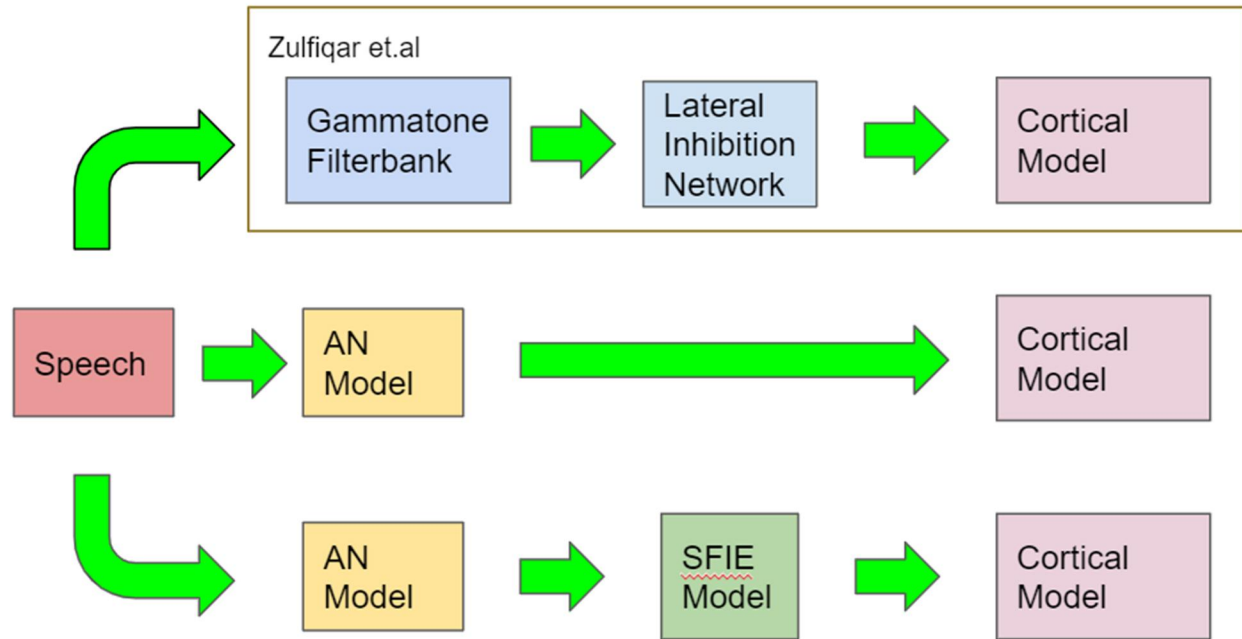


Figure 6. Experimental design schematic.

## 4.1 Vowel Contrast

Vowels are some of the easiest units of speech to understand computationally, due to their highly regular periodic structure, their ability to be characterized by spectral peaks, and their universal importance in human language. Phonetic analysis often plots vowels in terms of their first two formant frequencies (F1 and F2), peaks in the spectrum found to correspond to location of the articulators.

The vowels chosen to evaluate the models with different levels of complexity in the periphery were chosen such that they were highly contrasted in the vowel space. The five vowels used were /æ/, /ɛ/, /i/, /ʊ/, and /u/. Two 20-year-old native English speaking undergraduate students at the University of Rochester were recorded, and a center section of each vowel of about 150 ms was selected for analysis with the several models. These vowels were extracted from a wordlist of common English words, namely ‘bad’ (/bæd/), ‘bed’ (/bɛd/), ‘bead’ (/bid/), ‘buddhist’ (/budɪst/), and ‘boot’ (/but/).

Cortical output response to vowels was evaluated to explore representation of the pitch of the speaker (F0) and the first three formant frequencies.

## 4.2 Responses to Connected Speech

To best compare the changes in performance introduced by the use of a more physiologically realistic model for the auditory periphery, we aimed to replicate the analysis of Zulfiqar et al. of speech sounds. Sixty sentences from the TIMIT database were fed as input to the cortical model with the AN/IC periphery and the original gammatone/LIN periphery (Garofolo 1993). From a

normative perspective, a model of auditory cortex should encode as rich a set of linguistically important features (such as pitch of the stimulus and formants across time) in as efficient a manner as possible.

## 5 Results

### 5.1 Vowel Contrast

Five different vowels from only the male speaker of the two speakers were analyzed through different peripheral models to observe changes in frequency responses at the level of the cortex. Table I shows the first three formant frequencies for each of the vowels used.

Table 1. Formant Frequencies of Vowels Presented to Model

Vowel Phoneme	F1 (Hz)	F2 (Hz)	F3 (Hz)
æ	592	1576	2502
i	249	2305	2964
ε	469	1609	2527
u	295	952	2316
ʊ	352	1307	1723

Figure 7b shows the cortical output for each of the five vowels in all four model areas and for all three peripheral combinations. Circles of the same color as each vowel response are positioned at the formant frequencies. Important to note are the differences in rate for the vowels at each formant frequencies, and where peaks in the responses fall. The legend for all figure 7b graphs is shown below in Figure 7a.

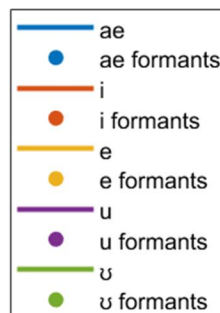
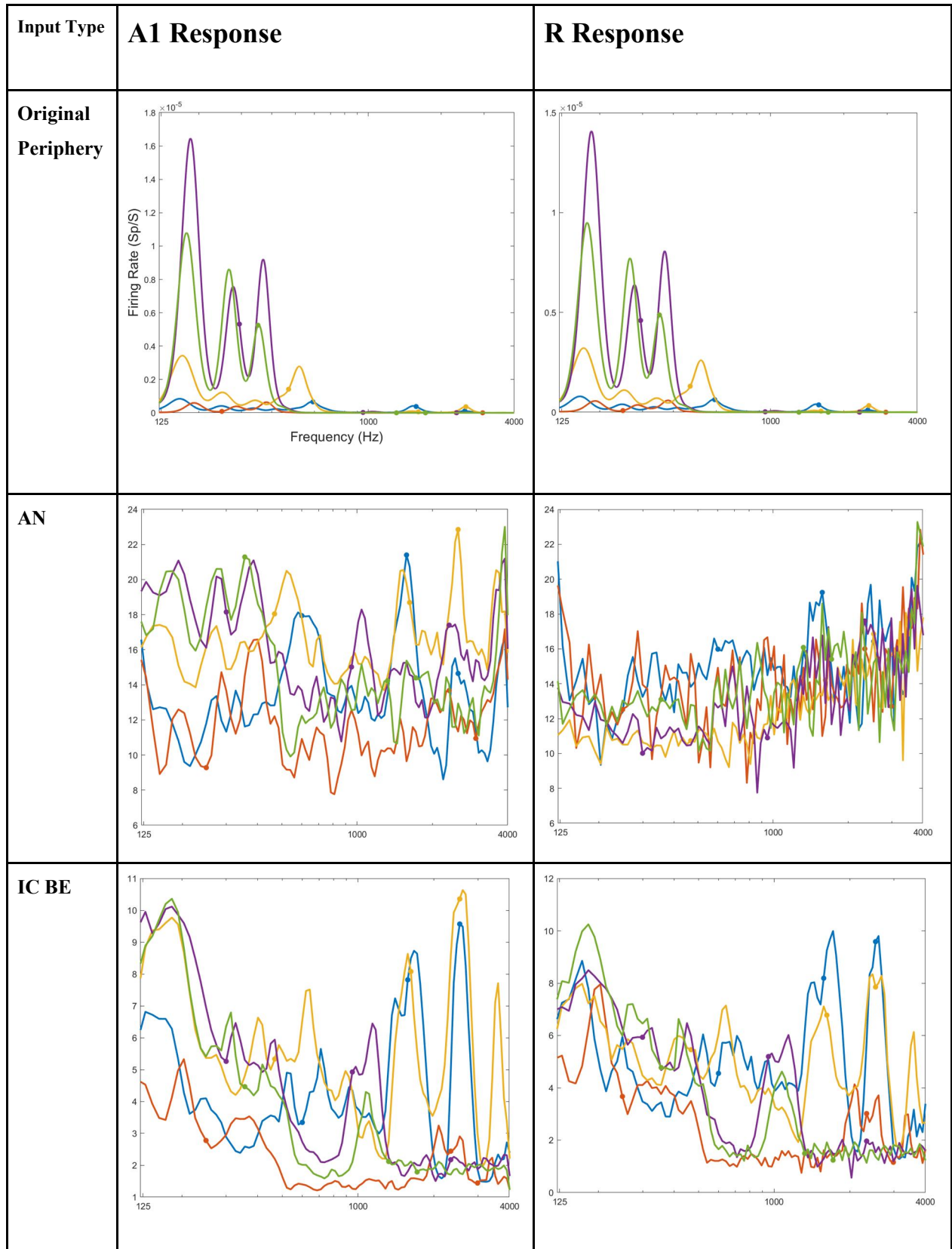


Figure 7a. Key of Formants for Figure 7b





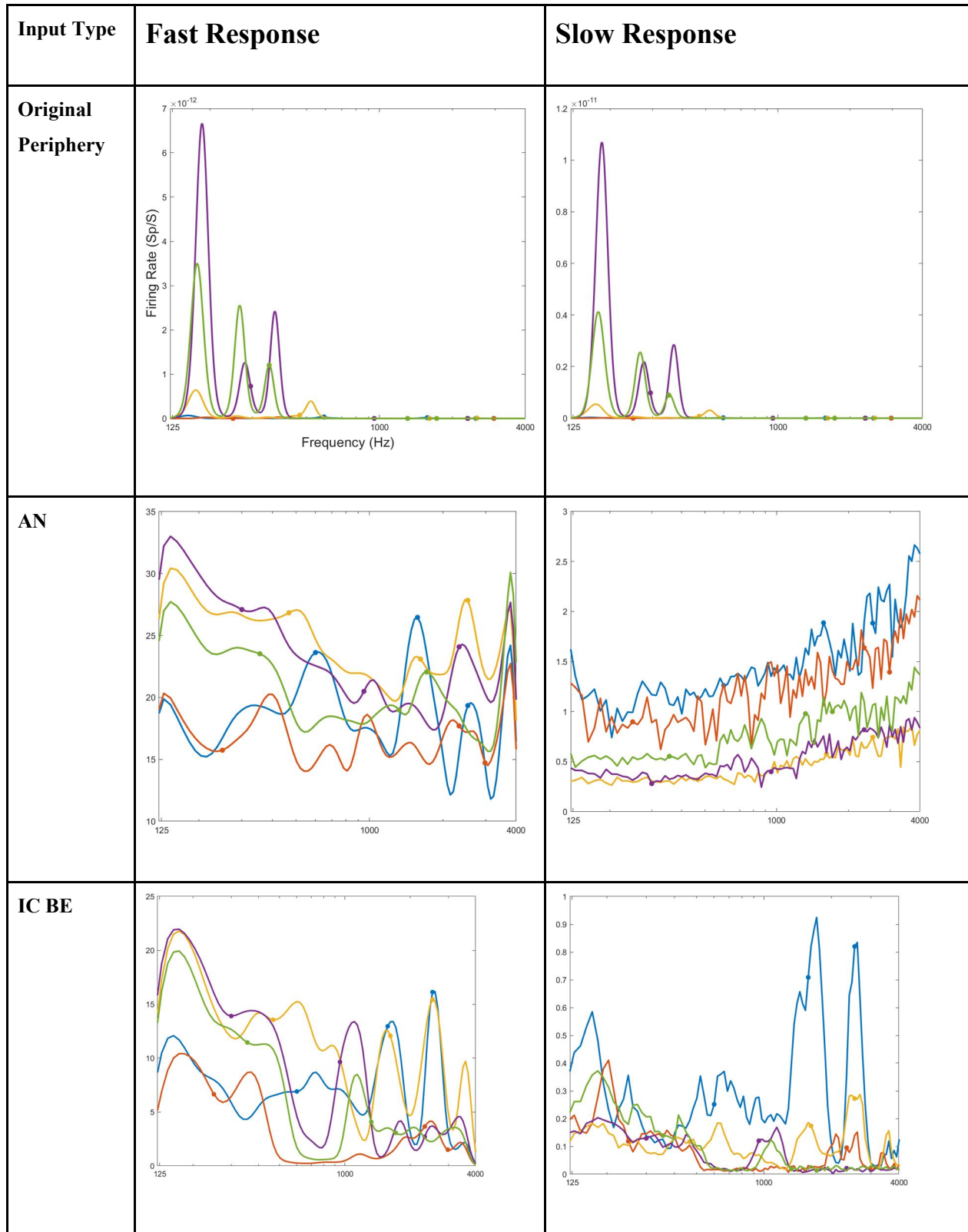


Figure 7b. Cortex model output for five spoken vowels. Outputs for three different peripheral models are shown for all four modeled areas of the auditory cortex. Dots on each curve represent the firing rate at the formant frequencies for that vowel.

Vowel contrast can be observed in Figure 7 by noting how far apart firing rates of model neurons are between vowels. For example, if ‘æ’ and ‘i’ have similar firing rates at a formant of ‘æ’ it is possible that the neural response between these two vowels cannot be distinguished at that formant. The human brain can tell these two apart very easily, so we expect vowel contrast to be high in the neural responses. It is also important to note that, since the Fast and Slow areas encode more abstract information than vowel identity, vowel contrast may not be represented in the same way in these areas than in the core (A1 and R) areas.

## **5.2 Responses to Sentences**

Results were quantified in a number of different methods. Mean spectrum coherence was calculated between the input and output to the cortical model, which summarizes how predictable the output would be by a linear least squares regression. A spectral encoding of pitch was also demonstrated across the four areas, especially the A1 and Slow areas. A sample response to a sentence is shown in Figure 8, with temporal and spectral profiles plotted on the corresponding axis. Note how the overall patterns of the responses are similar, but the modified periphery (the AN/IC model) preserve richer information in higher frequencies. Particularly in speech perception, we hypothesize that these higher frequencies are key

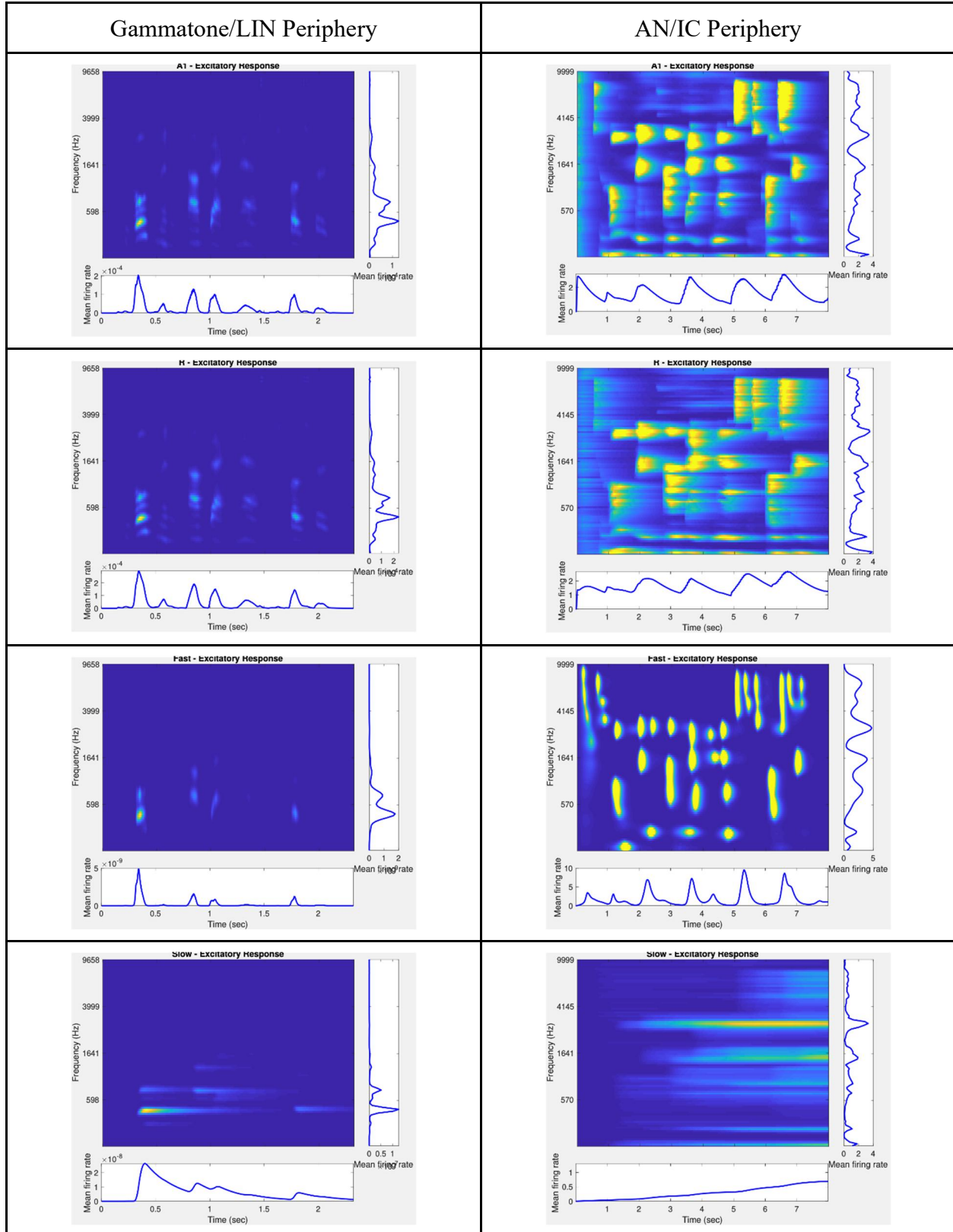


Figure 8: Example responses of each model to a sentence from the TIMIT database.

### 5.2.1 Mean Spectrum Coherence

Because there were no modifications made to the cortical stage of the model, one might expect mean spectrum coherence to be identical between the two models. However, there is a clear deterioration in all channels at later times in the results for the AN/IC periphery. Because the system does not change, this may imply that the type of input fed in by the gammatone/LIN periphery tends to evoke more linear responses than the AN/IC output from the same mode. Thus, these differences suggest spectrum coherence may speak more to the statistics of peripheral output than the actual computation of the cortical model. Thus, a gammatone/LIN periphery may be reducing the audio signal to a form which is not transformed as much by the computation of the cortical model, suggesting an oversimplification of the perceived sound.

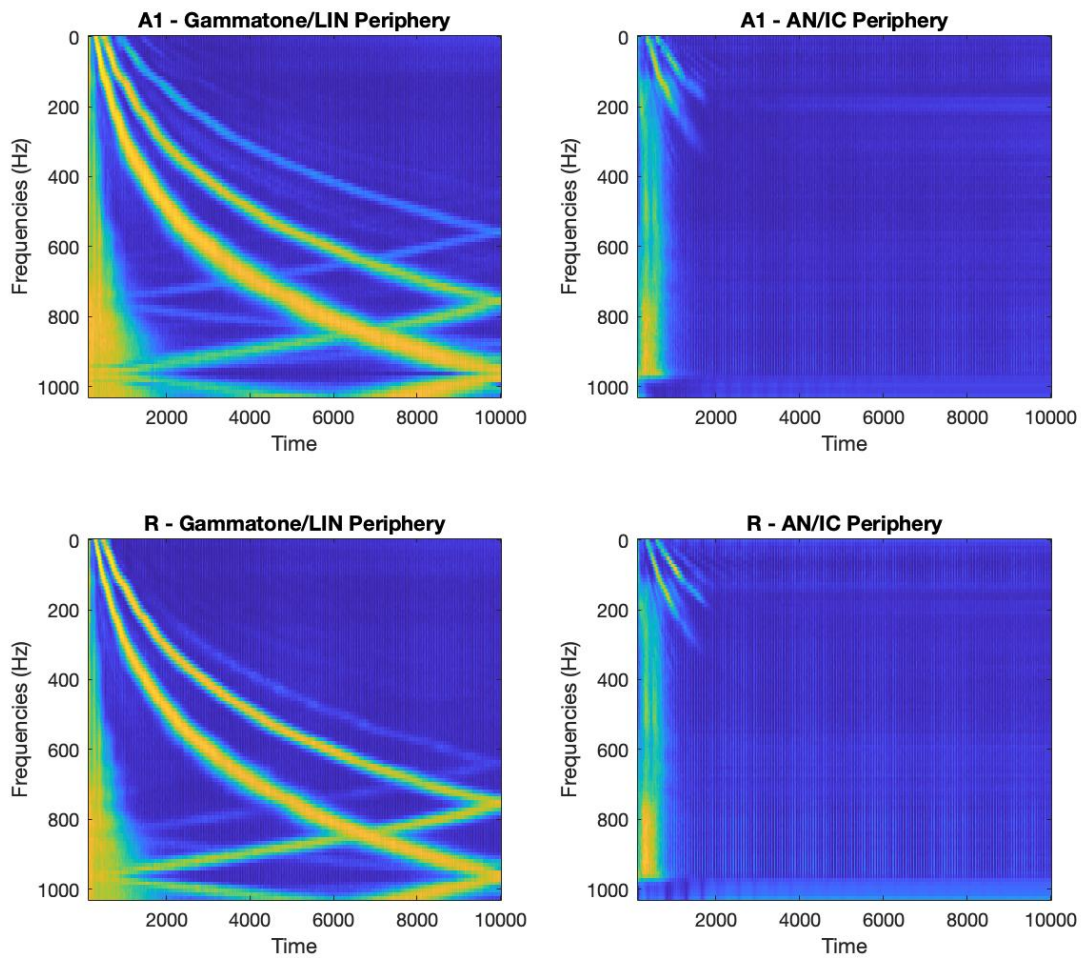


Figure 9: Mean spectrum coherence of output from peripheral and output from cortical model across all 60 sentences. Only results for excitatory populations of each area were computed. While these core areas show stark differences, these differences disappear in the subsequent belt areas of the models.

### 5.2.2 Pitch Estimation

One of the strongest results from Zulfiqar et al. was the dual encoding of pitch across the four areas; they observed that the mean spectral activity of the four regions mirrored the estimated fundamental frequency of the sentence, and that the mean temporal activity of the Fast area correlated strongly with the contour of the estimated pitch over time. Similar to their methodology, we used the YIN algorithm to obtain a pitch estimate for each sentence, an algorithm for estimating the fundamental frequency based off of the autocorrelation method, i.e. comparing subsections of a signal to find a match (Cheveigne and Kawahara 2002). The AN/IC model demonstrated a much clearer spectral encoding of this estimate in the four areas than the gammatone/LIN model, especially in the core areas (A1 and R). While the cortical model produced a local maximum near the estimated fundamental frequency, adding the AN/IC periphery made this spike much more obvious (Fig. 10). Furthermore, while using a linear periphery does represent the fundamental and first formant frequencies relatively well, it is easy to see that any higher formants cannot be easily distinguished in any of the areas, whereas using a more realistic periphery preserves the second and third formant quite well.

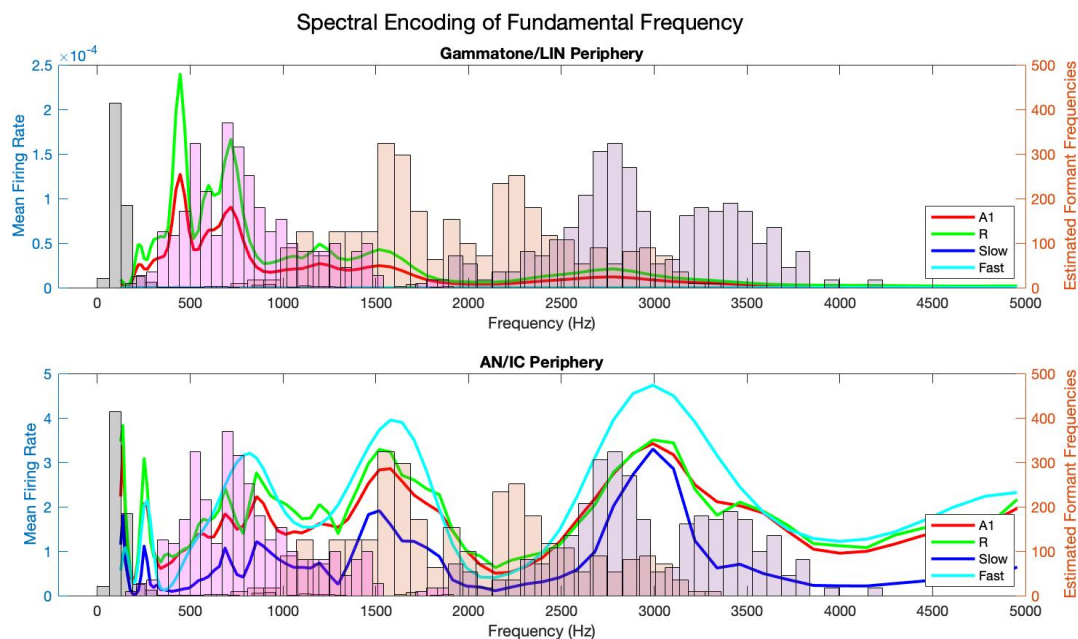


Figure 10. Spectral encoding of fundamental frequency of a sample sentence. Results were similar for all sentences tested. The gray histogram represents F0 estimates (as extracted by the YIN algorithm) and the magenta, orange, and purple histograms represent the first, second, and third formants respectively (as extracted by Praat).

## 6 Discussion

### 6.1 Vowel Contrast

An immediate observation from the vowel responses in Figure 7b is that much richer information about the frequency content of a sound is processed by the cortical model with a more realistic peripheral model. Past a frequency of about 1 kHz, there is very little frequency information encoded in the cortical response when a gammatone filterbank periphery is used. We know, however, that information is available to the brain at higher frequencies because we used formant frequencies for vowels above 1 kHz to differentiate and understand them. So, it is much more reasonable that speech drives neurons tuned to frequencies much higher than 1 kHz in the auditory cortex.

Furthermore, at formant frequencies, for both the AN and the AN+IC BE peripheries, there is a greater difference in firing rate between vowels than there is for the original periphery. One can imagine that a greater difference in neuron firing rate between speech sounds allows the brain to more easily distinguish between them. In this way, the peripheral models containing AN and AN+IC BE processing are more successful at differentiating between vowels, a skill that the human brain can accomplish with ease.

There are a few properties of the AN periphery that lead us to conclude that it is less appropriate for use in the cortical model than the IC BE cell output. The first is physiological. The output of the auditory nerve goes through an immense amount of processing before that information is received by the neurons in the auditory cortex. Next, the vowel responses show us that formant frequencies do not fall as often near peaks in the response for the AN periphery as they do for the AN+IC BE periphery. We would expect that the brain uses a change in firing rate near formant frequencies to identify vowels, and so a successful model will also show distinct features at formant frequencies. For example, peaks for F2 and F3 are similar for the vowels /ae/ and /e/ only in the IC BE periphery outputs. These vowels are similar in F2 and F3, and so this makes sense. In this way the IC BE input to the cortical model best shows realistic responses to vowels. The outputs for the IC BE periphery in A1 and R areas also show peaks at low frequencies, which could show improved pitch encoding; pitch encoding is examined more closely with regards to sentence responses.

The trends described above are most noticeable in the A1 and R outputs, which is consistent with the design of the cortical model, since these areas are the “core” areas that receive direct input from peripheral processing. The trends of formant frequencies and pitch are not as well seen in the Fast and Slow areas (the second table of Figure 7). This is reasonable, since Zulfiqar and colleagues posit that these areas encode fine temporal structure and fine-grained spectral properties of auditory stimuli, respectively (Zulfiqar et al. 2020).

A limitation of this analysis is that the responses shown are for only one speaker for one utterance, and only passed through the model once. The AN model incorporates internal noise (Nelson and Carney 2004), so any analysis is of course dependent on the summary statistics of a large dataset, rather than individual results.

Future analysis will include processing a larger dataset of vowel recordings and quantifying dispersion of the results from each model and then comparing how well each approach preserves the differences between vowels. Performance of vowel classifiers on the outputs of each model could also be used as a metric of distinguishability.

## 6.2 Sentence Response

The strongest singular result of our model in response to connected speech is the improved ability to represent higher formant frequencies. As shown in Figure 9, the spectral profile of the response in all four areas has a much richer shape, representing not just the fundamental frequency and first formant, but the second and third formant as well. From a linguistic perspective, this is vital, as even a simple, steady state vowel is determined by at least the first and second formant. Thus, it seems that embedding a more physiologically accurate model for the auditory periphery is necessary for linguistically relevant features of a sound to be encoded in this model of the auditory cortex. This model lends further support to the idea that the nonlinearities present in the computation of the inner ear and brainstem play a key role in linguistic perception.

## 7 Conclusion and Future Work

With this work we aimed to improve the responses of a computational model for auditory cortex to speech sounds by prepending a more physiologically realistic model for pre-cortical processing. We approach this “improvement” from a normative lens of preserving the right information rather than one of physiological replication, as we do not have neurophysiological with which to draw comparison. We observed a richer representation of higher frequencies in vowel responses, suggesting that an AN/IC periphery may better preserve the higher formants useful in distinguishing vowels. When presented with more complex stimuli, an AN/IC periphery achieved a clearer spectral encoding of estimated pitch of the sentence than a gammatone/LIN periphery did, one of the key results of the original paper proposing the cortical model used, and it quite clearly exhibited a spectral encoding of higher formant frequencies of the speech, a result that the original model was completely unable to match.

Future directions can examine other permutations of these same models for AN, IC, and AC. All experiments performed presented no inhibitory input to the cortical model. Variations of whether the band-enhanced or band-suppressed populations of the IC are used as excitatory or inhibitory input can be tested. The results presented here can be generalized by using a larger dataset and presenting summary statistics rather than sample outputs, but the principles that are exhibited in our results suggest that the use of physiologically accurate models for the auditory periphery are key to progress in the modeling of auditory cortex, especially in the context of speech perception.

## Acknowledgments

We would like to thank our advisors, Dr. Laurel Carney and Dr. Joyce McDonough at the University of Rochester for their guidance and support in all aspects of this paper. This work was performed to fulfill requirements for BME 216 - Speech on the Brain at the University of Rochester. We would also like to thank Isma Zulfiqar and the authors of “Spectro-Temporal Processing in a Two-Stream Computational Model of Auditory Cortex” for providing their model MATLAB code and support in using the code.



## References

- Arnott, S. R., M.A. Binns, C.L., Grady, and C. Alain. 2004. Assessing the auditory dual-pathway model in humans. *Neuroimage* 22, 401–408. doi: 10.1016/j.neuroimage.2004.01.014
- Belin, P., and R.J. Zatorre. 2000. “What”, “where” and “how” in auditory cortex. *Natural Neuroscience* 3, 965–966. doi: 10.1038/79890
- Carney, L.H. 1993. A model for the responses of low-frequency auditory-nerve fibers in cat. *JASA* 93:401–417
- Carney, L.H., et al. 2013. Detection thresholds for amplitude modulations of tones in budgerigar, rabbit, and human. *Advances in Experimental Medicine and Biology* 787: 391-8. doi:10.1007/978-1-4614-1590-9\_43
- Carney, L.H. and J.M. McDonough. 2019. Nonlinear auditory models yield new insights into representations of vowels. *Attention Percept Psychophys* 81:1034–1046. doi: 10.3758/s13414-018-01644-w
- Carney L.H., T. Li, and J.M. McDonough. 2015. Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *eNeuro*. doi: 10.1523/ENEURO.0004-15.2015
- Cheveigné, A.D. and H. Kawahara. 2002. YIN, a fundamental frequency estimator for speech and music. *JASA* 111.4:1917–1930. doi:10.1121/1.1458024.
- Delgutte, B. and N.Y. Kiang. 1984. Speech coding in the auditory nerve: I. Vowel-like sounds. *JASA* 75:866–878.
- Garofolo, J.S., et al. 1993. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium.
- Hillenbrand J.L., L.A. Getty, M.J. Clark, and K. Wheeler. 1995. Acoustic characteristics of American English vowels. *JASA* 97.5:3099-111.
- Jayakody, Dona. 2018. Impact of aging on the auditory system and related cognitive functions: A narrative review. *Frontiers in Neuroscience* 12. doi: 10.3389/fnins.2018.00125.
- Kaas, J.H. and T.A. Hackett, T. A. 2000. Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the U.S.A.* 97, 11793–11799. doi: 10.1073/pnas.97.22.11793
- Kaas, J.H., T.A. Hackett, and M.J. Tramo. 1999. Auditory processing in primate cerebral cortex. *Current Opinions in Neurobiology*. 9, 164–170.
- Ladefoged, P. 2012. *Consonants and Vowels*, 3rd edition. Wiley-Blackwell.
- Nelson P.C. and L.H. Carney. 2004. A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *JASA* 116:2173–2186
- Nelson, P.C. and L.H. Carney. 2007. Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *J. Neurophysiology* 91:522-539.
- Rauschecker, J.P. and B. Tian. 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences of the U.S.A.* 97:11800–11806. doi: 10.1073/pnas.97.22.11800
- Romanski, L. M., B. Tian, J., Fritz, M., Mishkin, P.S., Goldman-Rakic, and J.P. Rauschecker. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Natural Neuroscience* 2:1131–1136. doi:10.1038/16056
- Tian, B., D. Reser, A. Durham, A. Kustov, and J.P. Rauschecker. 2001. Functional specialization in rhesus monkey auditory cortex. *Science* 292:290–293. doi: 10.1126/science.1058911

- Wilson, H.R. and J.D. Cowan. 1972. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal* 12.1: 1-24.
- Wilson, H.R. and J.D. Cowan. 1973. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13:55–80.  
<https://doi.org/10.1007/BF00288786>
- Zilany M.S.A., I.C. Bruce, P.C. Nelson, and L.H. Carney. 2009. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *JASA* 126:2390–2412.
- Zulfiqar, I. et al. 2020. Spectro-Temporal Processing in a Two-Stream Computational Model of Auditory Cortex. *Frontiers in Computational Neuroscience* 13:95.  
[doi:10.3389/fncom.2019.00095](https://doi.org/10.3389/fncom.2019.00095)