CrossMark

Nonlinear auditory models yield new insights into representations of vowels

Laurel H. Carney [1] · Joyce M. McDonough [2]

Abstract

Studies of vowel systems regularly appeal to the need to understand how the auditory system encodes and processes the information in the acoustic signal. The goal of this study is to present computational models to address this need, and to use the models to illustrate responses to vowels at two levels of the auditory pathway. Many of the models previously used to study auditory representations of speech are based on linear filter banks simulating the tuning of the inner ear. These models do not incorporate key nonlinear response properties of the inner ear that influence responses at conversational-speech sound levels. These nonlinear properties shape neural representations in ways that are important for understanding responses in the central nervous system. The model for auditory-nerve (AN) fibers used here incorporates realistic nonlinear properties associated with the basilar membrane, inner hair cells (IHCs), and the IHC-AN synapse. These nonlinearities set up profiles of $f0$-related fluctuations that vary in amplitude across the population of frequency-tuned AN fibers. Amplitude fluctuations in AN responses are smallest near formant peaks and largest at frequencies between formants. These $f0$-related fluctuations strongly excite or suppress neurons in the auditory midbrain, the first level of the auditory pathway where tuning for low-frequency fluctuations in sounds occurs. Formant-related amplitude fluctuations provide representations of the vowel spectrum in discharge rates of midbrain neurons. These representations in the midbrain are robust across a wide range of sound levels, including the entire range of conversational-speech levels, and in the presence of realistic background noise levels.

Keywords Audition · Speech perception · Physiological psychology

The broad goal of this research is to understand the neural coding of the speech signal in the auditory pathway using computational neural models based on existing physiological data. We focus on the coding of vowel contrasts in the neural pathways from the auditory periphery to the midbrain. We present example model responses that illustrate how the properties of vowels that are essential for vowel contrasts, their formant structure, are robustly coded in the auditory pathway. Thus, the computational models presented here provide a framework for investigating the auditory system's contribution to the structure of vowel spaces.

Vowels and vowel spaces

The focus of the present study is to illustrate how the nonlinear responses of the auditory periphery influence the representation of vowels in the midbrain. Two principle reasons for focusing on vowels exist. First, the neural representations of vowel sounds differ substantially from acoustic representations, which have traditionally served as the basis for our understanding of vowels and vowel systems. Given a vowel sound, a critical question concerns the neural coding of that sound along the auditory pathway. Many auditory models used to explore speech responses use linear filter banks, which do not incorporate key nonlinear response properties of the auditory system. Yet the nonlinear response properties in the periphery shape the neural representations in ways that are

✉ Laurel H. Carney
Laurel.Carney@Rochester.edu

[1] Departments of Biomedical Engineering and Neuroscience, University of Rochester, 601 Elmwood Ave, Box 603, Rochester, NY 14642, USA

[2] Department of Linguistics, University of Rochester, Rochester, NY, USA

important for understanding more central neural responses. Second, vowel systems behave in systematic ways; they are composed of sets of phonemic contrasts that disperse themselves within an acoustically defined vowel space, in distinct cross-linguistic patterns, and independent of the number of vowels in any given system. The basis of these patterns is incompletely understood, and the effects of realistic physiological properties of the auditory system on these spaces is largely unexplored. The model described here provides a platform for future work investigating the possible constraints that the auditory system may impose on vowel spaces and on the structure of linguistic systems in general.

Attempts to model vowel spaces

How vowels pattern in linguistic systems has been under study for several decades, starting with work by Liljencrants and Lindblom (1972), with a hypothesis called dispersion theory (DT), later summarized as follows by Lindblom and Maddieson (1988, pp. 63): "If we know the number of vowels in an inventory, we know what their phonetic qualities are." The details underlying this observation relate to strong regularities in the way vowel systems develop and pattern cross linguistically. Languages disperse their vowels within an area defined by the first two vowel formants, irrespective of how many vowels are in the system (Crothers, 1978; Maddieson, 1984). Several investigators have built on this work, including, to name some classic examples, studies of the structure of consonant inventories (Lindblom & Maddieson, 1988), and revisions to Lindblom's DT, including "adaptive dispersion" and the inclusion of neural phase locking (Diehl & Kluender, 1989). The Grenoble group in the 1990s added a local focalization feature to dispersion (dispersion focalization theory, DFT) (Schwartz, Boë, Vallée, & Abry, 1997a, 1997b), to address the overpredictions along the F2 parameter. Stevens's (1972, 1989) quantal theory grounded the discussion in the asymmetries between articulatory movement and its acoustic output. Apart from quantal theory, the research on vowel spaces has had a strong auditory bias, based on the assumption that the objects of speech perception are auditory (Diehl & Kluender, 1989; Diehl, Lindblom, & Creeger, 2003; Kingston & Diehl, 1994; Nearey, 1997). Nearey (1997) laid out a "double-weak" hypothesis, which, while maintaining the strong auditory bias, proposed that three critical components interact with each other to shape the speech system: the independent production and perception systems and the abstract elements of the phonology. Becker-Kristal's (2010) UCLA dissertation took on a crucial issue underlying this discussion, the relationship between the symbolic system (phonemes) and the physical phonetic realization patterns, focusing on the phonetic patterns rather than the vowel symbols. Becker-Kristal's study addressed the vowel-space question

and dispersion-theory hypotheses based on an instrumental phonetic analysis of vowel systems using extensive data gathered from acoustic phonetic databases. Although these studies have been successful in addressing many aspects of the structure of vowel systems, unexplained differences between the predicted and actual vowel spaces persist.

The potential benefit of understanding the coding of speech sounds by the auditory system has long been recognized (e.g., Diehl, 2000; Diehl & Kluender, 1989; Diehl, Kluender, Walsh, & Parker, 1991; Diehl & Lindblom, 2004; Diehl et al., 2003; Liljencrants & Lindblom, 1972; Lindblom & Maddieson, 1988; Lindblom & Maddieson, 1988). While acoustic waveforms and spectrograms have been crucial to the study of speech for more than 100 years, the acoustic speech signal is the output of the production system and does not represent the responses of the auditory pathway nor the coding of the speech signal (Carlson & Granström, 1982; Diehl, 2008; Diehl et al., 2003; Lindblom, 1986; Nearey, 1997). The need to model the auditory responses have long been acknowledged, even in studies with a strong articulatory bias (Ghosh, Goldstein, & Narayanan, 2011). The role of temporal coding of sounds by phase-locked responses of auditory-nerve (AN) fibers has been explored (Diehl & Kluender, 1989; Diehl et al., 2003) with success in resolving some problems. But issues remain; these representations are not sufficient to explain the distribution of vowels in the F1–F2 space in the linguistic vowel systems. The set of vowels predicted by dispersion theory does not fully map the vowel spaces found in human languages, even when logarithmic frequency representations and frequency-dependent temporal information are considered (Diehl, 2008; Diehl et al., 2003) and even under the expansion of dispersion theory to include a local or focalization factor (Schwartz et al., 1997a, 1997b). Each step in this progression of studies has come closer to explaining the vowel space, but limitations remain, especially in the front-to-back (F2) and peripheral–nonperipheral vowel dimensions.

Modeling nonlinearities of the auditory periphery

The key nonlinearities that are included in the model of the auditory periphery used in this study are (i) cochlear compression and suppression in the mechanical responses of the basilar membrane, (ii) saturation of the transduction from mechanical to electrical responses in the sensory inner hair cells (IHC), and (iii) adaptation and saturation of the synapse between IHCs and AN fibers. Each of these stages of the model will be introduced here, and their effects on low-frequency fluctuations in AN responses will be described. Finally, the sensitivity of neurons in the auditory midbrain for low-frequency fluctuations, in the frequency range of voice pitch, will be described below. This report focuses on neural responses to vowel sounds, for which the relation between

fluctuation amplitudes and the harmonic spectrum are easily described. However, the concepts presented here rely upon profiles of low-frequency fluctuations that are present in responses to all complex sounds, including both voiced and unvoiced speech.

Cochlear compression and amplification

Compression is perhaps the most studied nonlinear property of the auditory periphery, partly because compression changes with sensorineural hearing loss. The cochlea can be described as an "active" amplifier of incoming sounds (Hudspeth, 2014; Kim, 1986). The amount of amplification is controlled locally along the frequency axis of the inner ear; that is, the gain of the so-called cochlear amplifier varies as a function of sound level at each place along the cochlea. The compressive nonlinearity refers to the decrease in cochlear amplification as sound level increases; the amplification is maximal (50–60 dB) at sound levels below 20–30 dB SPL and progressively decreases at higher levels. The sound level at one frequency influences the amplification not only at the place in the inner ear tuned to that frequency, but also across a small range of surrounding frequencies (Cody, 1992). The off-frequency influence on gain is referred to as suppression, which can "sharpen" the representation of sound spectra (Sachs & Young, 1980). Both compression and suppression involve level-dependent, and thus nonlinear, changes in the gain of the cochlear amplifier (Ruggero, Robles, & Rich, 1992). Cochlear amplification is also influenced by descending signals from the brain, via the auditory efferent system (Guinan, 2011), which receives inputs from both the periphery and from the midbrain (e.g., Terreros & Delano, 2015; reviewed in Carney, 2018). All of these factors that influence the cochlear amplifier play important roles in determining the amplitude of the mechanical input signal to the IHCs.

Inner-hair-cell saturation

The mechanical response of the cochlea is transduced into an electrical signal by the IHCs (see Fig. 1). The complete process of this transduction is an active area of study; transduction involves complex mechanical and fluid-coupled movements of microstructures in the inner ear (Howard, Roberts, & Hudspeth, 1988). Microelectrode recordings of the electrical signals in IHCs cells show that the electrical signal saturates as the input signal increases in amplitude (see Fig. 1). The act of making recordings from IHCs, which requires penetration of the cell membrane by an electrode, distorts the electrical signals (Zeddies & Siegel, 2004). Nevertheless, although the quantitative details of the input/output relationship are difficult to specify, reports agree that the electrical signal saturates gradually over a wide range of input sound levels (e.g., Dallos, 1985, 1986; Russell & Sellick, 1983; Russell, Richardson, & Cody,
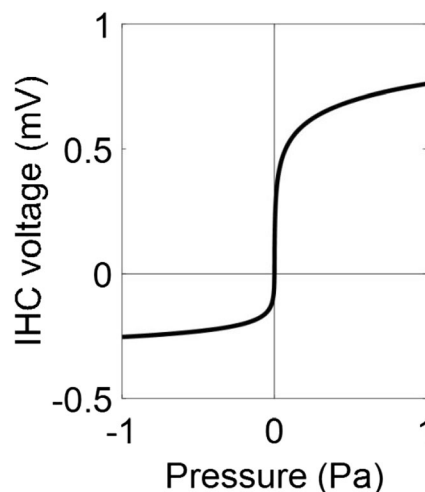

Fig. 1 Saturating input/output function that describes transduction from input pressure to output voltage in the IHC. As implemented by Zhang et al. (2001) and used in Zilany, Bruce, and Carney (2014)

1986). The description of IHC saturation in the AN model used here was based on in vivo recordings in the AN, which avoided damage to IHCs. The IHC stage of the AN model was designed to reproduce the effects of IHC saturation on AN responses (Zhang, Heinz, Bruce, & Carney, 2001).

IHC saturation becomes significant over the range of sound levels used for conversational speech (50–70 dB SPL), thus the influence of IHC saturation on responses to speech sounds is important to consider. The IHC saturating nonlinearity is often omitted from peripheral models for the following reason: It is a gradual saturation that occurs over a range of sound levels that is higher than the saturation of the IHC-AN synapse, described below. However, IHC saturation is critical in shaping the fluctuations in the time-varying responses of AN fibers and is particularly relevant for sounds at conversational speech levels (Carney, 2018).

"Capture" and its effect on fluctuations in AN responses

IHC saturation is associated with a key aspect of AN responses to harmonic sounds, referred to as synchrony capture (Deng, Geisler, & Greenberg, 1987). In response to harmonic sounds, AN-fiber responses are phase-locked to some combination of a harmonic near the fiber's characteristic frequency (CF) and to the beating between harmonics that creates a strong periodicity at $f0$. AN fibers tuned to a frequency near a peak in the spectral envelope, or formant, are dominated by the harmonic that is closest to the formant peak (see Fig. 2; Delgutte & Kiang, 1984; Miller, Schilling, Franck, & Young, 1997). This "capture" of the response by a single harmonic near CF can be explained by the combined actions of the cochlear amplifier and IHC saturation (see Fig. 1; Zilany & Bruce, 2007). Although capture is present in response to all
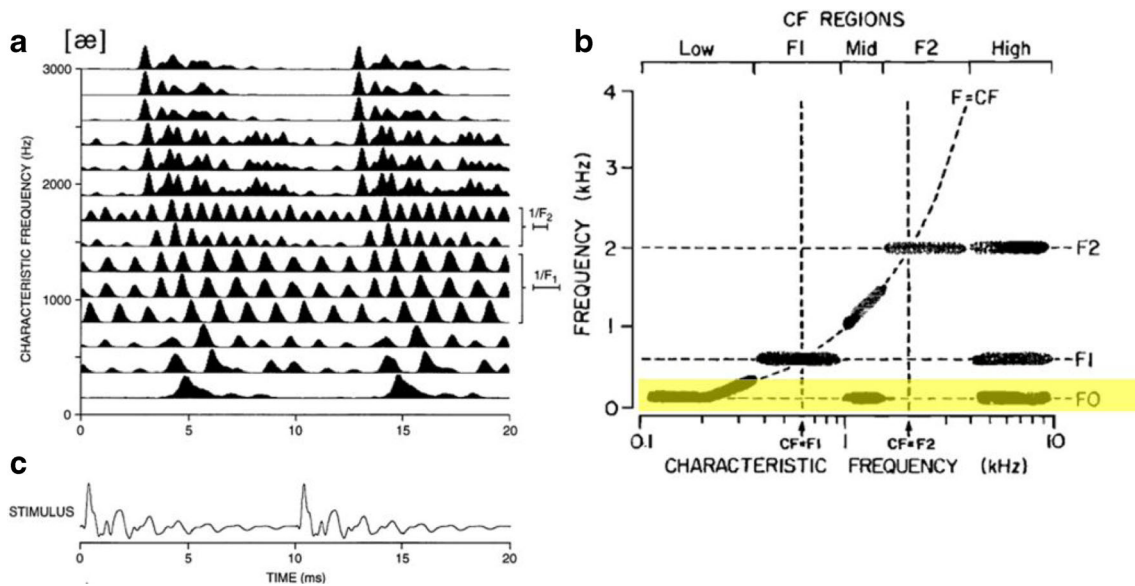
Fig. 2 Phase-locking to temporal fine structure and envelope features of vowels, and "capture" of neural timing by harmonics. The two forms of temporal information carried by AN fibers in response to a vowel are illustrated by peri-stimulus-time (PST) histograms (**a**) and by the dominant components, or the strongest periodicities in the AN responses, shown as a function of the AN fiber characteristic frequencies (CF) (**b**). Phase-locking to temporal fine structure near each fiber's CF (**a**) appears as frequency components in the AN responses at harmonic frequencies near formants (**b**, horizontal dashed lines) or near the fiber's CF (**b**, dashed curve). Temporal phase-locking to the vowel pitch, F0, is the result of the envelope fluctuations created by beating between two or more harmonics, observed in the AN responses as a strong periodicity locked to each pitch period (compare PST histograms in **a** to vowel waveform in **c**). Phase-locking to ƒ0 (**b**, highlighted rectangle) is reduced in fibers tuned near formant peaks (B, vertical dashed lines) due to synchrony capture, or dominance by a single harmonic near the spectral peak (see text). Synchrony capture is apparent on the left in the responses of fibers tuned near formants, which have responses dominated by a single harmonic and reduced phase-locking to the pitch period. **a** Modified from Delgutte (1987). **b** Modified from Delgutte and Kiang (1984)

harmonic sounds, this phenomenon is particularly strong in AN responses to vowels, for which vocal tract resonances shape the amplitudes of harmonics into peaks at the formant frequencies.

The nonlinear phenomenon of capture has generally been discussed in terms of its influence on phase-locking to temporal fine structure (Delgutte & Kiang, 1984; Deng et al., 1987; Miller et al., 1997). However, capture by one harmonic also reduces the influence on AN responses of ƒ0, which is due to beating between two or more harmonics. For example, the time-varying response of an AN fiber to a single harmonic has relatively weak amplitude fluctuations at ƒ0 (Fig. 2a; see responses of fibers with CF near F1 and F2), as compared with responses of fibers tuned between formants. Thus, the temporal response component at ƒ0 is reduced for AN fibers with CFs near formant peaks (Fig. 2b, highlighted rectangle). Because midbrain neurons are sensitive to amplitude fluctuations in the frequency range of voice pitch, capture plays a significant role in shaping the response of midbrain neurons to voiced sounds (see below).

Nonlinear response properties that are generally included in peripheral auditory models are the saturation of the IHC-AN synapse and the rectification of the signal (i.e., only "positive" discharges occur on AN fibers). All of the nonlinear response properties of the auditory periphery described

above, plus basic properties such as frequency tuning, are included in the computational AN model that was used to create the figures presented here (Zilany et al., 2014).

The temporal responses of AN fibers are characterized by both phase-locking to the detailed fine structure in waveforms and to the low-frequency fluctuations in amplitude associated with ƒ0 in voiced sounds (Joris, Schreiner, & Rees, 2004). The AN models used here include both of these types of temporal responses and their frequency dependence. The models for central auditory neurons include sensitivity to the temporal information in the peripheral responses. In particular, although the ability to phase-lock to temporal fine structure decreases along the ascending auditory pathway (Joris et al., 2004), many midbrain neurons are tuned to fluctuations in the frequency range of ƒ0. This sensitivity to a strong temporal aspect of voiced sounds provides another transformation of vowels sounds along the auditory pathway (Carney, Li, & McDonough, 2015).

It is worth emphasizing that the nonlinear properties described above, and included in the AN model, have substantial effects on neural responses to speech at the sound levels used for conversation. At these sound levels, single AN fibers typically respond to a very wide range of frequencies; these responses are not well described by models that focus on the sharp tuning of the inner ear that is the hallmark of threshold

tuning curves. Although it is beyond the scope of the present study, this computational AN model also allows "impairment" of the active cochlear amplifier and reduction of IHC sensitivity, two features of sensorineural hearing loss (Carney, Kim, & Kuwada, 2016). Here we will present responses of a model of the healthy periphery.[1]

Modeling the auditory midbrain

The auditory midbrain (inferior colliculus, IC, in mammals) is a major hub in the subcortical auditory system that receives convergent inputs from nearly all ascending auditory pathways. In terms of its position along the sensory processing pathway, the auditory midbrain is at the same level of sensory processing as the output neurons of the retina, the retinal ganglion cells. For example, both of these major sensory hubs project directly to the sensory thalamus; the auditory and visual thalamic regions that receive inputs from the auditory midbrain and retina then project to the primary cortical regions for each sensory system. Our understanding of neural representations of visual scenes is strongly influenced by an understanding of the complex properties of retinal output neurons. Thus, it is natural to reconsider auditory representations of speech sounds based on the response properties of auditory midbrain neurons.

The auditory midbrain is the first site along the ascending pathway where discharge rate depends strongly on amplitude

modulations in the stimulus (see Fig. 3). Similar to most auditory neurons, midbrain cells are tuned to a given audio frequency; this tuning is inherited from the frequency tuning of the inputs to the midbrain. Frequency tuning is first established by the tonotopic map in the cochlea. However, in addition to basic frequency tuning, most midbrain neurons have large (many fold) changes in discharge rate associated with tuning to the frequency of amplitude fluctuations in the stimulus. A simplified model of midbrain tuning to the frequency of input fluctuations is based on the interaction of excitatory and inhibitory inputs to the neurons (Carney et al., 2015; Nelson & Carney, 2004). This simple midbrain model omits several interesting features of midbrain neurons (e.g., binaural response properties) and focuses on the rate and timing of the responses to complex sounds that are presented to one ear.

Approximately 50% of midbrain neurons have band-enhanced tuning to fluctuation frequency, that is, the average response rate over some range of fluctuation frequencies is elevated with respect to the response to an unmodulated stimulus (see Fig. 3; Nelson & Carney, 2007). The other 50% have band-suppressed responses, with average rates over some range of fluctuation frequencies that are suppressed with respect to the response to unmodulated sounds. Band-enhanced neurons are generally broadly tuned bandpass filters, with best modulation frequency (BMF) typically in the range from 16–128 Hz (Langner & Schreiner, 1988; Nelson & Carney, 2007). Band-suppressed responses can be modeled as neurons that are excited by ascending inputs to the midbrain and inhibited by the band-enhanced neurons within the midbrain (see Fig. 4; Carney et al., 2015).

Band-enhanced neurons are driven well by fluctuations in the $f0$ range of most voices. Because synchrony capture reduces the $f0$ fluctuations in peripheral responses tuned near formant peaks, the responses of band-enhanced neurons tuned near formants are reduced (Carney et al., 2015). However, the band-enhanced neurons also tend to have weak responses in quiet, or if they are tuned to frequencies in a wide gap between formants (e.g., for midfrequency neurons responding to the vowel /i/). Thus, there is ambiguity in interpreting a weak response in the band-enhanced neurons—a weak response could indicate either a spectral peak that "flattens" the fluctuations of the inputs, or it could indicate a spectral valley. This ambiguity is resolved by band-suppressed neurons: these neurons respond strongly to spectral peaks that yield reduced fluctuations in peripheral responses, but they respond weakly in quiet or to inputs with large fluctuations (Carney et al., 2015). The response properties of band-suppressed midbrain neurons yield population responses that are more easily compared with the spectra of the stimuli; therefore, the population midbrain responses shown here will be based on models of band-suppressed neurons.

As described in Fig. 3, midbrain neurons have discharge rates that are sensitive to low-frequency ($f0$-related) temporal
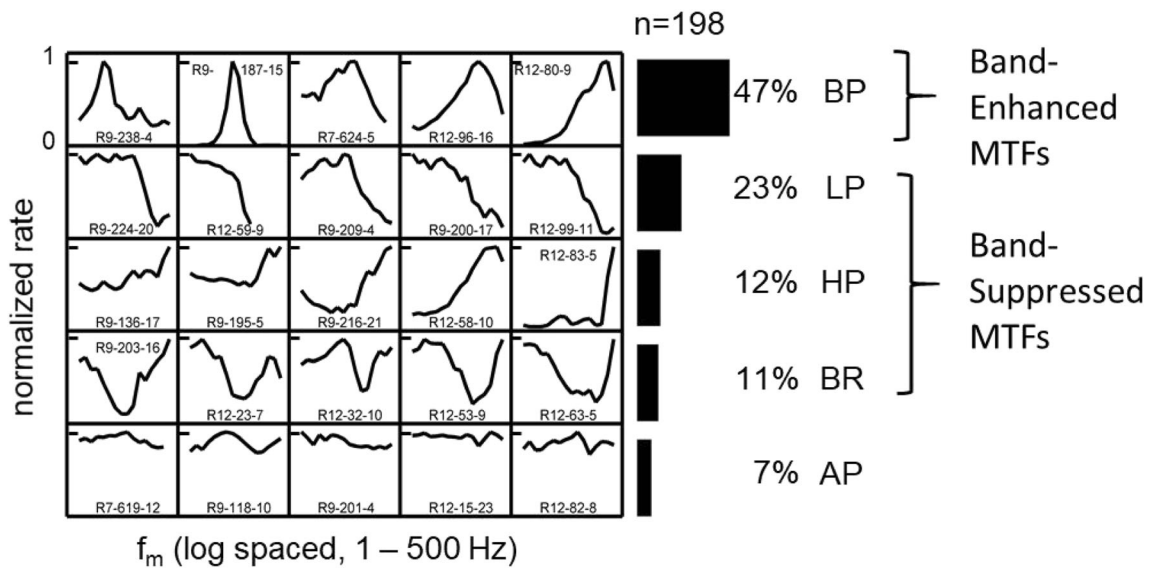
---

[1] Differences across AN fibers in synaptic saturation results in three different types of AN fibers: low-, medium-, and high -spontaneous -rate (LSR, MSR, HSR) fibers that differ in their thresholds and dynamic ranges (Liberman, 1978). The properties of these groups of AN fibers are often interpreted in terms of their average discharge rates; this simple approach suggests that saturated fibers cannot encode changes in the input signal. For this reason, several studies of neural coding have combined the three different types of AN fibers to explain the wide dynamic range of hearing (reviewed in Delgutte, 1996). This approach neglects the considerable information contained in the *time-varying* responses of AN fibers that have saturated *average* discharge rates. Here, we will focus on the responses of HSR fibers, the largest group of fibers, which have low thresholds and a relatively small dynamic range based on average discharge rate. HSR fibers provide the primary inputs to the ascending auditory pathway, and the fluctuation amplitudes of their time-varying responses encode speech spectra over the full range of levels encountered in speech sounds (Carney et al., 2015).

The LSR and MSR fibers have relatively wide dynamic ranges for average discharge rates, and thus they are often included in studies of neural encoding of complex sounds. Although these fibers project into the major ascending brainstem nuclei, the primary brainstem cells that project to the midbrain have low thresholds and limited dynamic ranges, consistent with dominance by their HSR inputs. Although we are not including them in the neural representations presented here, the LSR and MSR fibers provide exclusive inputs to a region of the brainstem that projects to the efferent system, which is involved in control of cochlear gain (reviewed in Carney, 2018). Thus, the LSR and MSR fibers may play a role in shaping the responses of HSR fibers, due to the important interaction of cochlear gain and IHC saturation. In addition to saturation and rectification of the IHC-AN synapse, adaptation of the synapse enhances responses to onsets in sounds, but also enhances synchronization to relatively low-frequency features in sound, including the pitch period, F0 (Zilany et al., 2009).

Fig. 3 Modulation transfer functions (MTFs) of midbrain neurons in cat illustrate the sensitivity of these cells' average discharge rates to amplitude fluctuations in a tone stimulus over a range of modulation frequencies (f_m). Stimuli were tones at each cell's characteristic frequency (CF), sinusoidally modulated across a range of low frequencies. Percentages of several different MTF types from one physiological study are shown (BP = band pass; LP = low pass; HP = high pass; BR = band reject; AP = all pass). Figure adapted from Nelson and Carney (2007)

fluctuations of their inputs. This sensitivity represents a translation of temporal information carried by AN responses into a profile of discharge rates across the population of midbrain neurons. Midbrain neurons also phase-lock to very low-
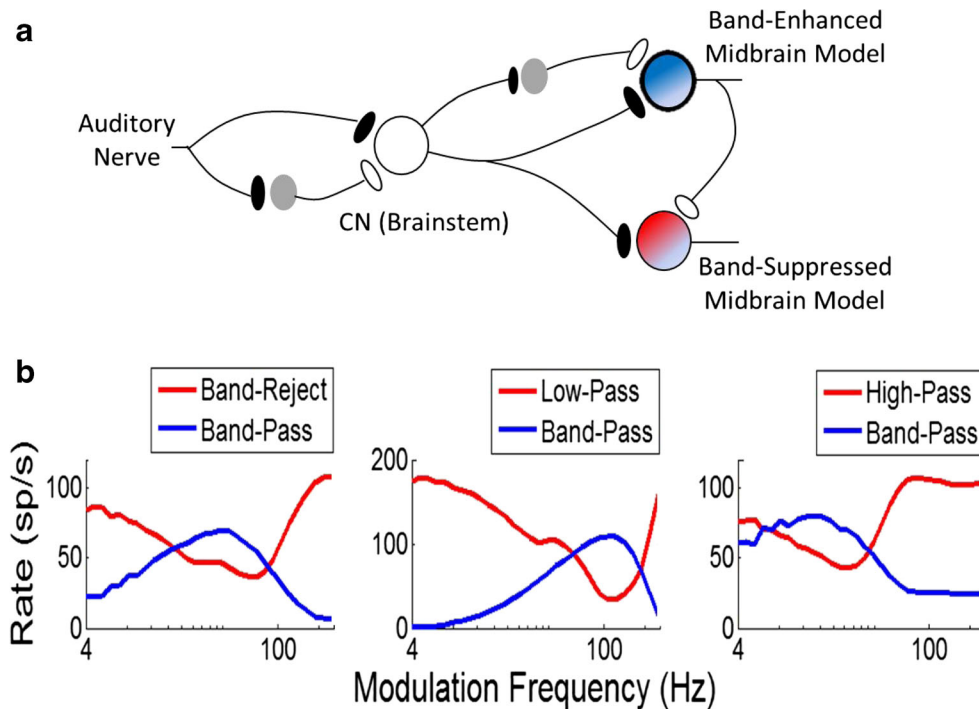

Fig. 4 Schematic diagram of midbrain neuron models and modulation transfer functions. **a** The SFIE model (blue; Nelson & Carney, 2004) is a simple combination of excitatory and inhibitory inputs, first at the level of the cochlear nucleus (CN) in the brainstem and again at the midbrain level. The band-suppressed model neuron (red) receives inhibition (white terminals) from the band-enhanced neuron, and excitation (black terminals) from the brainstem (Carney et al., 2015). **b** Most shapes of modulation transfer functions (MTFs) observed in the IC can be explained by these two simple models. The blue curves are band-enhanced MTFs; different best modulation frequencies (MTFs) result from different durations of the excitatory and inhibitory potentials in the model. The red curves illustrate different types of band-suppressed MTFs; these curves are suppressed with respect to the response to an unmodulated tone over some range of modulation frequencies. (After Carney et al., 2015)

frequency periodic inputs (including tones or the lower harmonics in a vowel) and to *f*0-related fluctuations (Joris et al., 2004). Investigation of the responses of fluctuation-sensitive midbrain neurons is therefore a logical step forward in the progression of studies of temporal coding of speech features (e.g., Diehl, 2008).

Method

Vowel stimuli

Vowel waveforms were from the Hillenbrand, Getty, Clark, and Wheeler (1995) database. The figures below show responses to a subset of eight speakers from the database. Selected speakers had average *f*0s that were approximately evenly spaced, ranging from 95 to 223 Hz (Speaker IDs: M03, M23, M08, M33, M40, W46, M10, W38). Illustrations below show responses to the vowels /α/ (see

Fig. 5), and /i/, /e/, /æ/, and /u/ (see Figs. 6, 7, and 8). For each speaker, the average *f*0 across all vowels was used to specify the best modulation frequency of the midbrain model used to process that speaker's vowel waveforms.

The formant frequencies of the vowels in the Hillenbrand et al. (1995) database change over time (Hillenbrand & Nearey, 1999). The model neural responses track the changes in the formants during the course of the vowel (see Rao and Carney, 2014); however, the changes in the formant frequencies tend to degrade the response profiles which are based on average responses over fixed time windows. To reduce the effect of changes in the formant frequencies on the average model responses, the central 100 ms of the steady-state portion of each vowel was extracted, and 5-ms raised-cosine on/off ramps were applied. Responses were studied for the vowel recordings provided in the Hillenbrand et al. (1995) database or to the same sounds with added background noise. The added noise had a long-term spectrum that matched the average speech spectrum (LTASS noise; Byrne et al., 1994). To
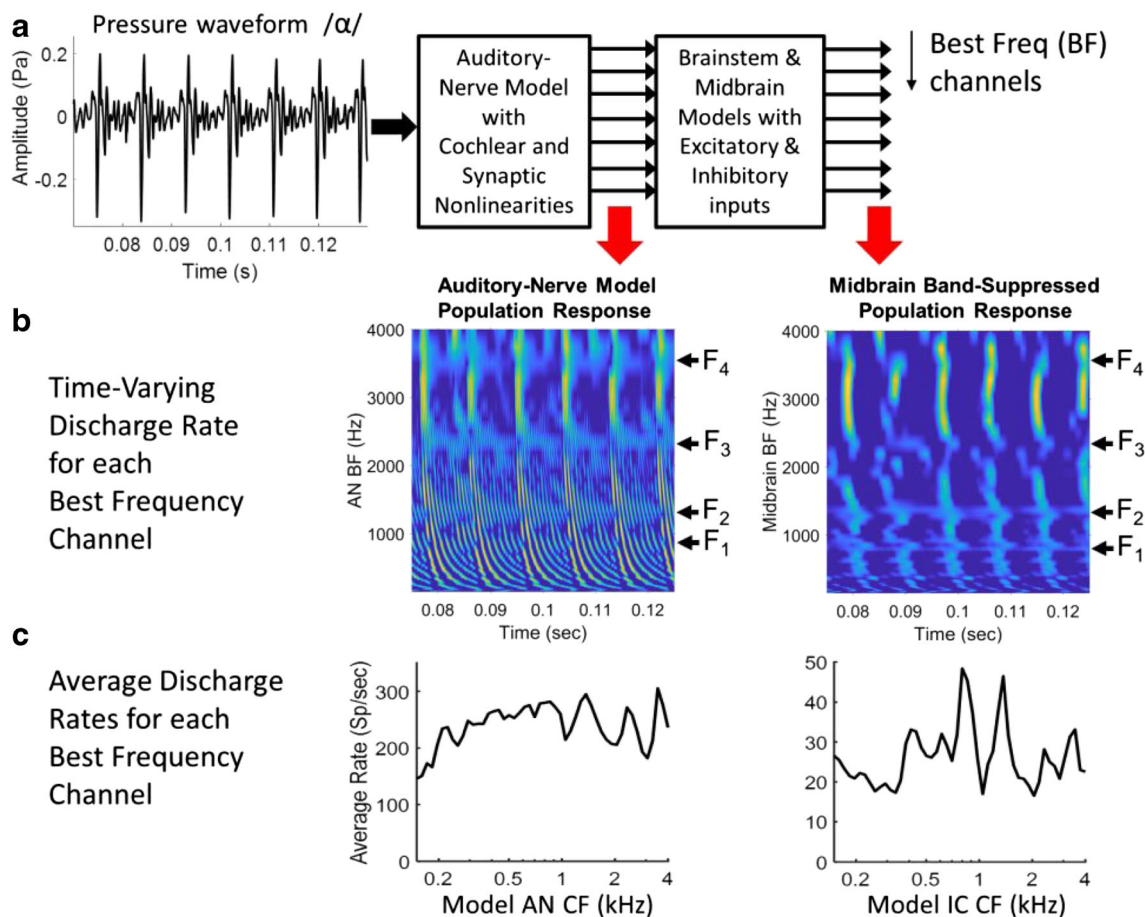

Fig. 5 a Schematic diagram showing stimulus waveform, AN population model, and brainstem/midbrain population models. The stimulus is /α/ from Hillenbrand et al. (1995), spoken by a male with average *f*0 = 128 Hz, presented at 65 dB SPL. Formant frequencies: F_1 = 748 Hz, F_2 = 1293 Hz, F_3 = 2446 Hz, F_4 = 3383 Hz. **b** Time-frequency population responses. All AN model fibers are high-spontaneous-rate fibers; 50 BF channels from 150 to 4000 Hz. Midbrain responses are for band-suppressed neurons created by band-enhanced cells with BMF = 128 Hz (see Fig. 4). **c** Discharge rates averaged over time, for each BF channel, plotted on logarithmic frequency axes; model AN responses (left) and midbrain responses (right)
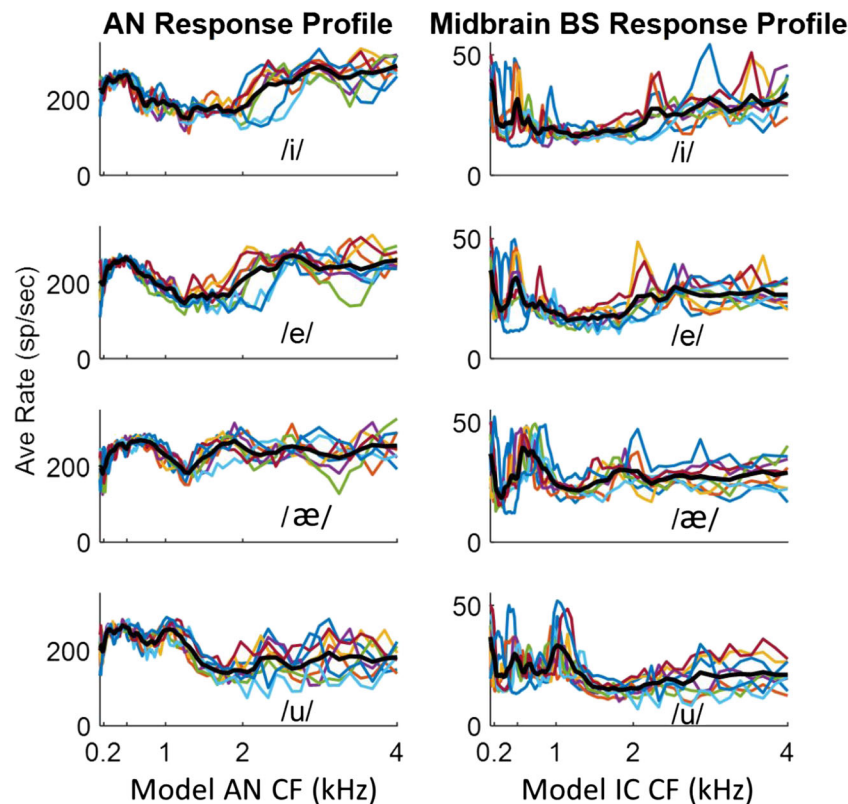
Fig. 6 Model AN and midbrain responses to four vowels (/i/, /e/, /æ/, and /u/, from top to bottom), spoken by eight speakers in the Hillenbrand et al. (1995) database. All vowels were scaled to 70 dB SPL. The left column shows population average discharge rate profiles for model AN responses, and the right column shows band-suppressed (BS) model midbrain responses. Each midbrain model had a BMF that was matched to the average $f0$ of each speaker across the vowels in the Hillenbrand et al. (1995) database. Averages rate responses across the eight speakers are shown as thick black lines in each panel

adjust the signal-to-noise ratio (SNR) of the speech in background noise, the speech level was kept constant (70 dB SPL for Figs. 6, 7, and 8), and the noise level was set at 65 dB SPL overall level, for a +5 dB SNR (see Figs. 7 and 8).

AN and IC model implementations

The illustrations here are based on models for AN fibers and for neurons in the auditory midbrain (inferior colliculus, IC, in mammals) that have band-suppressed responses to amplitude modulations. The AN model (Zilany et al., 2014) has the key nonlinearities mentioned above: cochlear compression/suppression, IHC saturation, and synaptic adaptation and saturation. The model was developed based on physiological responses of the AN in cat (Carney, 1993, Zhang et al., 2001; Zilany & Bruce, 2006, 2007; Zilany, Bruce, & Nelson, 2009). The version of the AN model used here has sharpened tuning estimated from physiological and psychophysical measures (Shera, Guinan, & Oxenham, 2002), as appropriate for the human ear (Ibrahim & Bruce, 2010).

The midbrain model used here is based on the band-suppressed model in Carney et al. (2015), modified to allow convenient adjustment of the BMF of band-enhanced neurons.

The BMF of model neurons was approximately matched to the average $f0$ of each speaker. The band-suppressed responses, shown here, were excited by the same ascending inputs as the band-enhanced neurons, and were inhibited by the band-enhanced model neuron (see Fig. 4). The model parameters and their dependence on BMF are described in Table 1.

The time-varying rate function, which represents the probability of neural responses as a function of time, for each model stage, $R_{output}(t)$, in response to an input rate function, $R_{input}(t)$, is described by

$$R_{output} = A \left[\{ \alpha(\tau_{ex},t) * R_{input}(t) \} - C_{inh} \{ \alpha(\tau_{inh},t) * R_{input}(t - D_{inh}) \} \right],$$

where $\alpha(\tau)$ is the alpha function used to simulate excitatory and inhibitory response potentials, described by $\alpha(\tau,t) = t\, e^{-t/\tau}$. The symbol * represents the convolution operation. The output of each stage was half-wave rectified. For speed, the alpha functions were implemented in the frequency domain; the scalars in Table 1 were applied to alpha functions that were normalized to have an area equal to 1.

Both physiological and psychophysical estimates of midbrain modulation transfer functions suggest that they are broad, with quality (Q = CF/bandwidth) factors equal to approximately 1 (i.e., the bandwidth is equal to the peak frequency.) Responses
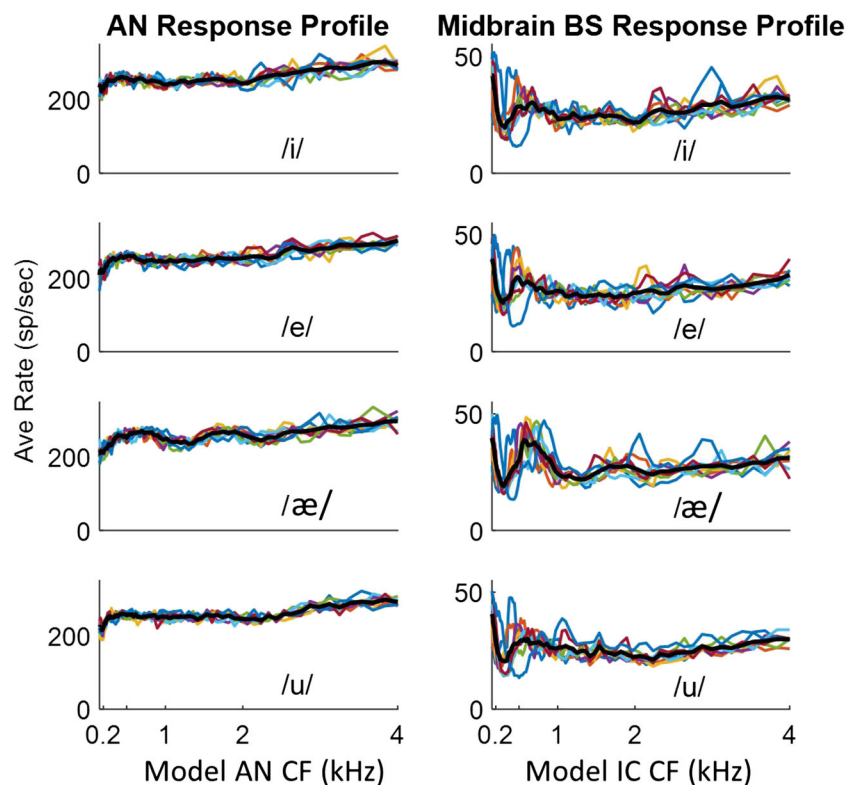
Fig. 7 Model AN and midbrain band-suppressed rate profiles for vowels at +5 dB SNR, in a background of LTASS noise. (Vowels were presented at 70 dB SPL, and the added LTASS noise was at 65 dB SPL). Note that at this moderate noise level, the AN responses are largely saturated, but the model midbrain response profiles (bold black lines in the right-hand column) still have peaks at many of the formant frequencies

of band-suppressed model neurons are shown here; the speech features represented in these responses also influence the responses of band-enhanced neurons, except that the representations are "inverted." The presence of two groups of midbrain cells that represent stimulus features with opposite "polarities" is reminiscent of retinal ganglion cells with on-center-off-surround and off-center-on-surround receptive fields. Recall that auditory midbrains are comparable, in terms of sensory processing level, to retinal ganglion cells: Both cell groups project directly to the sensory thalamus.

The response profiles shown below (see Figs. 6, 7, and 8) are based on 80 frequency channels, log-spaced from 150 to 4000 Hz. The AN model includes a time-varying spontaneous rate, which introduces variability in the simulations from repetition to repetition, and across different model neurons in the population. For the population responses presented here, 10 independent AN fibers were simulated in each frequency channel, and averages across these fibers were used as inputs to the midbrain models. This number of AN fibers is a conservative estimate of the number of AN fibers that are present in each frequency channel, as defined by the number of AN fibers that innervate each inner hair cell in the cochlea, which ranges from approximately eight to 30 (Keithley & Schreiber, 1987).

The computer code used to generate the illustrations presented here is available at https://www.urmc.rochester.edu/labs/carney.aspx.

Results

A schematic diagram illustrating how the AN and midbrain models were used to create population profiles is shown in Fig. 5. The pressure waveform for the approximately steady-state portion of the vowel /ɑ/ was extracted (from the word /hod/) for speaker M08 in the Hillenbrand et al. (1995) database. The central 100 ms of the vowel was used as input to the model; a 50-ms segment is illustrated in Fig. 5a–b for clarity. The AN response is shown in Fig. 5 for 50 HSR fibers with best frequencies (BFs) tuned from 150 to 4000 Hz, spaced evenly on a log frequency scale. Figure 5b shows the time-varying rate of each AN fiber over 50 ms of the vowel response. The two types of temporal responses described above are apparent in these responses: phase-locking to the fine structure is clearest in responses to the low BFs, and phase-locking to *f*0 is clear across all BFs (the six vertical bands in the population response are responses across all BFs to the onset of each pitch period.).

The effect of synchrony capture can be observed in the time-varying AN responses in Fig. 5b (left). For AN fibers with BFs near F1, F2, and even near F3, the responses are dominated by phase-locking to the harmonic that is closest to the formant frequency. These responses show up as "smooth" bands that have peaks evenly spaced in time, at the frequency of the dominant harmonic. The phase-locking

Responses to 70 dB SPL Vowels in Quiet   **Responses to 70 dB SPL Vowels in 65 dB SPL Noise**
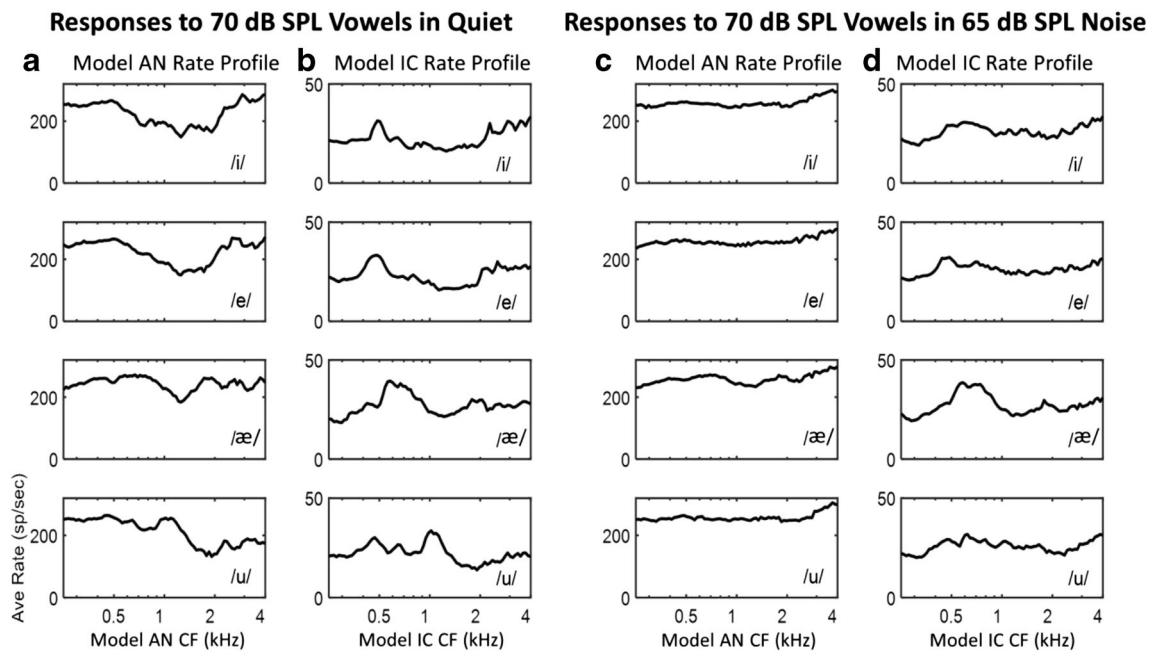

Fig. 8 Model response profiles for vowels in quiet (**a, b**) and in +5 dB SNR added noise (**c, d**). Model AN (**a, c**) and band-suppressed midbrain (**b, d**) profiles for the four vowels in Figs. 6 and 7, averaged across the eight speakers

of these "captured" fibers to *f*0 is relatively weak, as compared to the fibers with BFs between the formant frequencies.

The band-suppressed model midbrain neurons (see Fig. 5b, right) respond most strongly to inputs that have relatively weak fluctuations; thus, these responses are strongest near the formants, especially near F1 and F2. The changes in responses near F3 and F4 are not as clear in the time-frequency plots of Fig. 5b. However, the changes in the response profile across all frequency channels are more apparent in the plots of average discharge rates (see Fig. 5c). For these plots, 10 independent AN fibers were simulated for each frequency channel. The average rate across the entire 100 ms segment of the steady-state vowel, averaged across the 10 model AN fibers, is plotted as a function of BF (Fig. 5c, left). The average discharge rates are also shown for the population of band-suppressed model midbrain neurons (Fig. 5c, right). Each model midbrain neuron received an input based on an average of the ten independent AN simulations.

The frequencies of F1–F4 can be observed in the average discharge rate profile of the AN fibers, visible as four peaks at the appropriate frequencies. These peaks in the AN model rates are susceptible to added noise, and will vary from speaker to speaker, as shown below. However, the *f*0-related temporal information that is embedded in the AN responses (Fig. 5b, left) results in an enhanced representation of the formants, especially F1 and F2, in the rate profile across the population of model midbrain neurons (Fig. 5c, right). It should be noted that *f*0 is also encoded in the temporal responses of the model midbrain neurons.

The strategy used to compute the model rate profiles in Fig. 5c were applied to four other vowels in Figs. 6, 7, and 8.

Responses for these figures were computed for eight speakers with a range of average *f*0s. For each speaker, a midbrain model with BMF matched to the average *f*0 for that speaker across all vowels in the Hillenbrand et al. (1995) database was used. For each response profile, the AN responses were averaged across ten independent simulations for each of 80 frequency channels, evenly spaced on a log axis from 150 to 4000 Hz. The average of the ten AN simulations was provided as the excitatory input to the midbrain models (see Fig. 4); only the band-suppressed (BS) midbrain model responses are shown in Figs. 6, 7, and 8.

Figure 6 shows responses of AN and midbrain models to four vowels presented in "quiet" (i.e., no additional noise was added to the database recordings.) The colored lines represent each of the eight speakers. The thick black line is an average of the model responses across the eight speakers. The formant locations are visible in the AN rate profiles (Fig. 6, left column), even in the responses of these HSR model AN fibers, which have nearly saturated average discharge rates at the 70 dB SPL presentation level used. As observed in Fig. 5, the representations of formants are enhanced in the model midbrain responses (Fig. 6, right column), due to the temporal properties of the model AN responses.

A challenge to any neural representation based on average discharge rates is the addition of background noise. Human listeners with normal hearing can communicate at relatively low signal-to-noise ratios (SNRs), though many neural models (and automatic speech recognition systems) cannot. For the responses in Fig. 7, vowel levels were held fixed at 70 dB SPL, and 65 dB SPL LTASS noise was added, for a +5

Table 1 Parameters for the modified same-frequency inhibition-excitation (SFIE) model used for the simulations presented here

	Excitatory input time constant τ_{ex} (ms)	Inhibitory input time constant τ_{inh} (ms)	Inhibitory delay D_{inh} (ms)	Amplitude of inhibition re: excitation C_{inh}	Scalar for model stage output A
Brainstem model	0.5	2.0	1.0	0.6	1.5
Midbrain Band-enhanced (BE) model	$\tau_{exBE} = 0.1/BMF$	$\tau_{inhBE} = 1.5\ \tau_{exBE}$	$2\ \tau_{exBE}$	0.9	1
Midbrain Band-suppressed (BS) model	$\tau_{exBS} = 0.1/BMF$	$\tau_{inhBS} = \tau_{inhBE}$	1	4	0.5

Note. The model is modified from that in Nelson and Carney (2004) and Carney et al. (2015) to allow convenient adjustment of the parameters based on a desired best modulation frequency (BMF) for the band-enhanced midbrain model

dB SNR. At this SNR, the AN response profiles are largely flat, as has been observed in other studies of AN coding (reviewed in Delgutte, 1996). However, the temporal responses, especially phase-locking to *f*0, is still relatively robust in these AN responses, as reflected in the model midbrain responses, which still encode many of the formant frequencies based on peaks in the rate profiles.

Figure 8 compares the AN and midbrain responses, averaged across the eight speakers, for both quiet and +5 dB SNR vowel stimuli. Comparison of the panels in Fig. 8a and c illustrate the deleterious effect of the added noise on the representations of formants in terms of peaks in the average rate profiles. However, the temporal aspects of the model AN responses that are important for driving the model midbrain responses are relatively robust, even in the presence of noise. Comparison of the plots in Fig. 8b (in quiet) and d (in noise) illustrate that peaks in the midbrain profiles related to formant frequencies are generally still present, although the peaks are diminished in noise with respect to those in quiet. Measures of intelligibility predicted by these responses would be degraded, but not eliminated, by the added noise. It should be noted that these plots only show representations of the vowel responses based on average rate. The temporal response properties of the midbrain neurons convey additional information related to both the formant locations and *f*0. This temporal information in midbrain responses is important in explaining formant discrimination thresholds in a noise background in the parakeet, an animal model that has discrimination thresholds comparable to human listeners (Henry et al., 2017). Future studies of the representations of vowels using these computational models should include both average rate and temporal response metrics.

Discussion

The goal of this paper is to introduce computational modeling tools for auditory responses in the periphery and midbrain, and to illustrate how peripheral auditory nonlinearities and midbrain response properties may inform questions about the organization of linguistic speech patterns. Examples of responses to vowels, in quiet and in background noise, demonstrate the potential utility of these tools for exploring neural representations of speech. Model response profiles (see Figs. 6, 7, and 8) include the effects of several nonlinear properties in the periphery. By including the sensitivity of midbrain neurons to low-frequency fluctuations of their neural inputs, profiles are also shaped by temporal responses in the periphery to the combination of the fine structure and *f*0-related fluctuations in vowels. Any realistic description of auditory representations of speech must be robust at conversational speech levels (shown here) and across a range of sound levels (Carney et al., 2015). Furthermore, the representations must be robust in background noise levels for which speech communication is possible. Here, the focus was on introducing these models and testing these basic requirements. These computational tools provide a tool for ongoing tests of hypotheses concerning the structure of vowel spaces.

The response profiles of model AN and midbrain neurons (e.g., Figs. 6, 7, and 8) can be used to explore hypotheses related to the structure of the vowel space, such as dispersion theory and dispersion focalization theory. For example, the model rate profiles can be used to assess "distances" between different vowels, using the strategy of Lindblom (1986), which was inspired by Plomp's (1970) quantification of the similarity of sounds with timbre. Note that these representations, and the distances between them, are affected by the vowel spectra, as well as by the temporal information in AN fibers. This temporal information gradually transitions from a combination of phase-locking to both fine structure and envelope (*f*0-related) fluctuations in low-frequency channels, to phase-locking to *f*0alone in high-frequency channels where phase-locking to fine structure is diminished. A reevaluation of distances between vowels and correlations to perceptual similarities can be pursued using the neural representations provided by the models presented here.

Becker-Kristal (2010) examined the acoustic realization patterns of the phonemic contrasts in vowel systems to test the veracity of Lindblom's dispersion theory, and the dispersion and focalization theory (DFT) of the Schwartz group (Schwartz et al., 1997a, 1997b). These theories hypothesize that inventories are structured to maximize perceptual

contrasts among the vowels, with the focalization feature favoring a local merging of close vowel formants. The Becker-Kristal study was based on acoustic measurements of phonemic representations of vowels, drawn from published data from 550 studies, resulting in 304 separate inventories. His results largely corroborated dispersion theory with noteworthy differences. Among them were that the vowels within a system disperse themselves in an acoustic space defined by the number of vowels in the system, with larger inventories in larger acoustic spaces, with significant differences between the behavior of peripheral and nonperipheral vowels. The increases in the number of peripheral vowels affect the F1 space. He observes that inventories with increasing numbers of nonperipherals tend to prefer vowels front of the midline, with a F2 high close to F3, a focalization feature, and to disprefer retracted vowels, building vertical systems and overriding the repulsion of peripheral vowels in the system.

Becker-Kristal (2010) demonstrated that the vowel space is well defined by an acoustic space and auditory principles (Diehl & Kluender, 1989; Diehl et al., 2003). The model discussed here demonstrates that the spectral envelope of vowels is robustly encoded in midbrain response profiles. Thus, changes in temporal coding by auditory neurons as a function of frequency, together with the cochlear frequency map, shape the responses in a manner that can be investigated in the context of dispersion theory.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and dispersion theory: Insights from a large cross-linguistic corpus.* Unpublished dissertation, University of California, Los Angeles.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., … Kiessling, J. (1994). An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America*, 96(4), 2108–2120.

Carlson, R., & Granström, B. (1982). Towards an auditory spectrograph. In R. Carlson & B. Granström (Eds.), The representation of speech in the peripheral auditory system (pp. 109–114), Amsterdam: Elsevier.

Carney, L. H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat. *The Journal of the Acoustical Society of America*, 93(1), 401–417.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: Implications for sensorineural and hidden hearing loss. *Journal of the Association for Research in Otolaryngology, 19*(4), 331–352.

Carney, L. H., Kim, D. O., & Kuwada, S. (2016). Speech coding in the midbrain: Effects of sensorineural hearing loss. In P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, & C. Lanting (Eds.), Physiology, psychoacoustics and cognition in normal and impaired hearing (pp. 427–435). New York: Springer.

Carney, L. H., Li, T., & McDonough, J. M. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Eneuro, 2*, ENEURO-0004.

Cody, A. R. (1992). Acoustic lesions in the mammalian cochlea: Implications for the spatial distribution of the 'active process'. *Hearing Research*, 62(2), 166–172.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), Universals of human language (Vol. 2), pp. 99–152). Stanford: Stanford University Press.

Dallos, P. (1985). Response characteristics of mammalian cochlear hair cells. *Journal of Neuroscience, 5,* 1591–1608.

Dallos, P. (1986). Neurobiology of cochlear inner and outer hair cells: Intracellular recordings. *Hearing Research, 22,* 185–198.

Delgutte, B. (1987). Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination. In M. E. Schouten (Ed.), The psychophysics of speech perception (pp 333–353). Amsterdam: Springer.

Delgutte, B. (1996). Physiological models for basic auditory percepts. In H. L. Hawkins, T. A. McMullen, & R. R. Fay (Eds.), Auditory computation (pp. 157–220). New York: Springer.

Delgutte, B., & Kiang, N. Y. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *The Journal of the Acoustical Society of America, 75,* 866–878.

Deng, L., Geisler, C. D., & Greenberg, S. (1987). Responses of auditory-nerve fibers to multiple-tone complexes. *The Journal of the Acoustical Society of America, 82,* 1989–2000.

Diehl, R. (2000). Searching for an auditory description of vowel categories. *Phonetica*, 57, 267–274.

Diehl, R., Kluender, K., Walsh, M. & Parker, E. (1991). Auditory enhancement in speech perception and phonology. In R. Hoffman & D. Palermo (Eds.), Cognition and the symbolic processes, Vol 3: Applied and ecological perspectives (pp. 59–76). Hillsdale: Erlbaum.

Diehl, R., & Lindblom, B. (2004). Explaining the structure of feature and phoneme inventories. In S. Greenberg, W. Ainsworth, A. Popper, R. Fay (Eds.), Speech processing in the auditory system (101–162). New York: Springer-Verlag.

Diehl, R., Lindblom, B., & Creeger, C. (2003). Increasing realism of auditory representations yields further insights into vowel phonetics. *Causal Publications Adelaide* (Vol. 2, pp. 1381–1384).

Diehl, R. L. (2008). Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society London B: Biological Science, 363,* 965–978.

Diehl, R. L., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121–144.

Ghosh, P. K., Goldstein, L. M., & Narayanan, S. S. (2011). Processing speech signal using auditory-like filterbank provides least uncertainty about articulatory gestures. *The Journal of the Acoustical Society of America, 129*(6), 4014–4022. doi:https://doi.org/10.1121/1.3573987

Guinan, J. J. (2011). Physiology of the medial and lateral olivocochlear systems. In D. K. Ryugo, R. R. Fay, & A. N. Popper (Eds.), Auditory and vestibular efferents (pp. 39–81). New York: Springer.

Henry, K. S., Abrams, K. S., Forst, J., Mender, M. J., Neilans, E. G., Idrobo, F., & Carney, L. H. (2017). Midbrain synchrony to envelope structure supports behavioral sensitivity to single-formant vowel-

like sounds in noise. *Journal of the Association for Research in Otolaryngology, 18,* 165–181.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099–3111.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America, 105,* 3509–3523.

Howard, J., Roberts, W. M., & Hudspeth, A. J. (1988). Mechanoelectrical transduction by hair cells. *Annual Review of Biophysics and Biophysical Chemistry, 17*(1), 99–124.

Hudspeth, A. J. (2014). Integrating the active process of hair cells with cochlear function. *Nature Reviews Neuroscience, 15*(9), 600.

Ibrahim, R. A., & Bruce, I. C. (2010). Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In E. Lopez-Poveda, A. Palmer, & R. Meddis (Eds.), The neurophysiological bases of auditory perception (pp. 429–438). New York: Springer.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84,* 541–577.

Keithley, E. M., & Schreiber, R. C. (1987). Frequency map of the spiral ganglion in the cat. *The Journal of the Acoustical Society of America*, *81*(4), 1036–1042.

Kim, D. O. (1986). Active and nonlinear cochlear biomechanics and the role of outer-hair-cell subsystem in the mammalian auditory system. *Hearing Research*, *22*(1/3), 105–114.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70,* 419–454.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat: I. Neuronal mechanisms. *Journal of Neurophysiology, 60,* 1799–1822.

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. *The Journal of the Acoustical Society of America, 63,* 442–455.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48,* 839–862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), Experimental phonology (pp. 13–44). Orlando: Academic Press.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. N. Li (Eds.), Language, speech and mind: Studies in honor of Victoria Fromkin (pp. 62–78). London: Routledge.

Maddieson, I. (1984). Patterns of sounds (Cambridge Studies in Speech Science and Communication) Cambridge: Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511753459.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers. *The Journal of the Acoustical Society of America, 101,* 3602–3616.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America, 101*(6), 3241–3256.

Nelson, P. C., & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America 116*, 2173–2186.

Nelson, P. C., & Carney, L. H. (2007). Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *Journal of Neurophysiology, 97,* 522–539.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), Frequency analysis and periodicity detection in hearing (pp. 397–414). Sijthoff, Leiden.

Rao, A., & Carney, L. H. (2014). Speech enhancement for listeners with hearing loss based on a model for vowel coding in the auditory midbrain. *IEEE Transactions on Bio-Medical Engineering, 61*(7), 2081–2091.

Ruggero, M. A., Robles, L., & Rich, N. C. (1992). Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression. *Journal of Neurophysiology*, *68*(4), 1087–1099.

Russell, I. J., Richardson, G. P., & Cody, A. R. (1986). Mechanosensitivity of mammalian auditory hair cells in vitro. *Nature, 321,* 517–519.

Russell, I. J., & Sellick, P. M. (1983). Low-frequency characteristics of intracellularly recorded receptor potentials in guinea-pig cochlear hair cells. *The Journal of Physiology, 338, 179*–206.

Sachs, M. B., & Young, E. D. (1980). Effects of nonlinearities on speech encoding in the auditory nerve, *The Journal of the Acoustical Society of America, 68,* 858–875.

Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997a). Major trends in vowel system inventories. *J. Phonetics* 25: 233-253.

Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997b). The dispersion-focalization theory of vowel systems. *Journal of Phonetics, 25,* 255–286.

Shera, C. A., Guinan, J. J., & Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements *Proceedings of the National Academy of Sciences*, *99*(5), 3318–3323.

Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3–46.

Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory–acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view (pp. 51–66). New York: McGraw-Hill.

Terreros, G., & Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Frontiers in Systems Neuroscience, 9*(134), 1–8.

Zeddies, D. G., & Siegel, J. H. (2004). A biophysical model of an inner hair cell. *The Journal of the Acoustical Society of America*, *116*(1), 426–441.

Zhang, X., Heinz, M. G., Bruce, I. C., & Carney, L. H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *The Journal of the Acoustical Society of America*, *109*(2), 648-670.

Zilany, M. S., & Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *The Journal of the Acoustical Society of America*, *120*(3), 1446–1466.

Zilany, M. S., Bruce, I. C., Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America, 135,* 283–286.

Zilany, M.S., Bruce, I.C., Nelson, P.C., Carney, L.H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. The Journal of the Acoustical Society of America 126:2390-2412.

Zilany, M. S. A., & Bruce, I. C. (2007). Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats. *The Journal of the Acoustical Society of America, 122,* 402–417.