

---

*Research Article: Theory/New Concepts | Sensory and Motor Systems*

## **Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations**

Speech Coding in the Brain

**Laurel H. Carney<sup>1</sup>, Tianhao Li<sup>1,\*</sup> and Joyce M. McDonough<sup>2</sup>**

<sup>1</sup>*Departments of Biomedical Engineering and Neurobiology & Anatomy, University of Rochester, 601 Elmwood Ave, Box 603, Rochester, NY 14642, USA*

<sup>2</sup>*Department of Linguistics, University of Rochester, 503 Lattimore Hall, PO BOX 270096, Rochester, NY 14627-0096 USA*

DOI: 10.1523/ENEURO.0004-15.2015

Received: 13 January 2015

Revised: 18 June 2015

Accepted: 18 June 2015

Published: 2 July 2015

---

**Author Contributions:** LHC, TL and JMM developed the model and wrote the paper; LHC and TL performed the physiological experiments and analyzed the data.

**Funding:** NIH-NIDCD: 001641.

The authors declare no competing financial interests.

Funding: NIH-NIDCD (001641)

\*Current Address: Department of Electrical and Control Engineering, Liaoning Technical University, HuLuDao, LiaoNing 123000, China

Corresponding author: Laurel H. Carney, Departments of Biomedical Engineering and Neurobiology & Anatomy, University of Rochester, 601 Elmwood Ave, Box 603, Rochester, NY 14642 USA. E-mail: [Laurel.Carney@Rochester.edu](mailto:Laurel.Carney@Rochester.edu)

**Cite as:** eNeuro 2015; 10.1523/ENEURO.0004-15.2015

**Alerts:** Sign up at [eneuro.org/alerts](http://eneuro.org/alerts) to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

eNeuro

<http://eneuro.msubmit.net>

eN-TNC-0004-15R2

Speech Coding in the Brain: Representation of Vowel Formants by Midbrain  
Neurons Tuned to Sound Fluctuations

1 Speech Coding in the Brain:

2 Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations

3

4 Abbreviated Title: Speech Coding in the Brain

5 Laurel H. Carney<sup>1</sup>, Tianhao Li<sup>1\*</sup>, Joyce M. McDonough<sup>2</sup>

6 <sup>1</sup>Departments of Biomedical Engineering and Neurobiology & Anatomy, University of

7 Rochester, 601 Elmwood Ave, Box 603, Rochester, NY 14642 USA.

8 <sup>2</sup>Department of Linguistics, University of Rochester, 503 Lattimore Hall, PO BOX 270096

9 Rochester, NY 14627-0096 USA.

10 \*Current Address: Department of Electrical and Control Engineering, Liaoning Technical

11 University, HuLuDao, LiaoNing 123000, China

12 Author Contributions: LHC, TL and JMM developed the model and wrote the paper; LHC and

13 TL performed the physiological experiments and analyzed the data.

14 Corresponding author:

15 Laurel H. Carney

16 Departments of Biomedical Engineering and Neurobiology & Anatomy

17 University of Rochester

18 601 Elmwood Ave, Box 603

19 Rochester, NY 14642 USA.

20 E-mail: Laurel.Carney@Rochester.edu

21

22 Number of Figures: 9

23 Number of Tables: 1

24 Abstract: 246 words

25 Introduction: 750 words

26 Discussion: 1101 words

27

28 Acknowledgements: This work was supported by NIH-NIDCD R01-001641. Drs. David A.

29 Cameron and Kenneth S. Henry provided valuable comments on the manuscript.

30 The authors declare no competing financial interests.

31 **Abstract**

32           Current models for neural coding of vowels are typically based on linear descriptions of  
33 the auditory periphery and fail at high sound levels and in background noise. These models rely  
34 on either auditory-nerve (AN) discharge rates or phase-locking to temporal fine-structure.  
35 However, both discharge rates and phase-locking saturate at moderate to high sound levels, and  
36 phase-locking is degraded in the central nervous system at mid to high frequencies. The fact that  
37 speech intelligibility is robust over a wide range of sound levels is problematic for codes that  
38 deteriorate as level increases. Additionally, a successful neural code must function for speech in  
39 background noise at levels that are tolerated by listeners. The model presented here resolves  
40 these problems and incorporates several key response properties of the nonlinear auditory  
41 periphery, including saturation, synchrony capture, and phase-locking to both fine-structure and  
42 envelope temporal features. The model also includes the properties of the auditory midbrain,  
43 where discharge rates are tuned to amplitude fluctuation rates. The nonlinear peripheral response  
44 features create contrasts in the amplitudes of low-frequency neural rate fluctuations across the  
45 population. These patterns of fluctuations result in a response profile in the midbrain that  
46 encodes vowel formants over a wide range of levels and in background noise. The hypothesized  
47 code is supported by electrophysiological recordings from the inferior colliculus of awake rabbit.  
48 This model provides information for understanding the structure of cross-linguistic vowel spaces  
49 and suggests strategies for automatic formant detection and speech enhancement for listeners  
50 with hearing loss.

51

52

53 **Significance Statement**

54           Encoding of speech sounds is the most important function of the human auditory system.  
55    Current models for neural coding of speech fail over the range of sound levels encountered in  
56    daily life and in background noise. The acoustic structure of vowels and the properties of  
57    auditory midbrain neurons that are tuned to low-frequency amplitude fluctuations suggest a  
58    neural code for the spectral peaks (called formants) that identify vowels. The proposed neural  
59    code for speech sounds is the first that is robust over a wide range of sound levels and in  
60    background noise. These results address classic problems in auditory neuroscience and  
61    linguistics and suggest novel strategies for auditory prosthetics, automatic speech recognition,  
62    and speech enhancement for hearing aids and telephones.

63

64 **Introduction**

65 Vowels carry a heavy functional load in all languages, especially in running speech and  
66 discourse. How vowels are encoded by the nervous system across the range of sound levels  
67 important for vocal communication is unknown. The acoustic signature of vowels includes  
68 periodicity at the fundamental frequency (F0, or voice pitch), harmonics of F0, and formants, the  
69 amplitude bands in the spectrum that characterize vowel contrasts (Fant, 1960). The first two  
70 formants are most important for vowel identification. Studies of auditory-nerve (AN) speech  
71 coding typically focus on response rates or temporal synchrony at frequencies to which a fiber is  
72 most sensitive (Delgutte and Kiang, 1984; Sachs and Young, 1979; Schilling et al., 1998; Young  
73 and Sachs, 1979). These codes are adequate for low-level speech sounds in quiet, but they fail  
74 for moderate to high sound levels and in background noise. Vowels also induce systematic  
75 changes in the amplitude of F0-related fluctuations in AN responses. The vowel-coding  
76 hypothesis tested here focuses on the F0-related neural fluctuations and on contrasts in their  
77 amplitudes across neurons tuned to different frequencies.

78 Many inferior colliculus (IC) neurons display both spectral tuning, described by a most  
79 sensitive best frequency (BF), and tuning to the frequency of sinusoidal fluctuations in  
80 amplitude, described by a best modulation frequency (BMF) (Joris et al., 2004; Krishna and  
81 Semple, 2000; Nelson and Carney, 2007). Most IC neurons tuned for amplitude fluctuations have  
82 BMFs in the range of voice pitch (Langner, 1992) and are thus well suited to represent the  
83 critical acoustic features of vowels (Delgutte et al., 1998). The vowel-coding hypothesis  
84 presented here takes advantage of nonlinear properties of AN responses, including rate saturation  
85 (Sachs and Abbas, 1974; Yates et al., 1990) and synchrony capture, which is the dominance of a  
86 single stimulus frequency component on the response (Fig. 1; Deng and Geisler, 1987; Miller et

87 al., 1997; Young and Sachs, 1979). These nonlinearities have strong effects on rate fluctuations  
88 of AN fibers in response to vowels and provide a robust framework for encoding vowel features.

89 Figure 1 introduces the main features of the vowel-coding hypothesis. The harmonic  
90 structure of vowels (Fig. 1A) yields two types of periodicity that co-exist in AN responses:  
91 phase-locking to harmonics near the fiber's BF, and phase-locking to slow fluctuations at F0.  
92 Fibers tuned near formant peaks may be saturated, and these fibers also have sustained responses  
93 that are dominated by phase-locking to a single harmonic near the fiber's BF, referred to as  
94 synchrony capture. Both of these nonlinearities result in responses with relatively weak  
95 fluctuations at F0 (Fig. 1B). Fibers tuned to frequencies away from formants are not dominated  
96 by one harmonic but are influenced by the beating of multiple harmonics, resulting in strong  
97 low-frequency neural fluctuations at F0 (Fig. 1C).

98 The contrast in amplitude of low-frequency rate fluctuations across the AN population is  
99 enhanced in the midbrain by the rate tuning of IC neurons to amplitude modulations, described  
100 by modulation transfer functions (MTFs) (Fig. 1D,E). The majority of MTFs in the IC have  
101 band-pass tuning to amplitude modulations (Fig. 1D), and the rest have band-reject (Fig. 1E),  
102 low- or high-pass tuning (Nelson and Carney, 2007), or more complex MTFs that combine  
103 excitation and inhibition (Krishna and Semple, 2000). Midbrain cells with band-pass MTFs that  
104 have *maxima* (i.e. best modulation frequencies, BMFs) near F0 are hypothesized to display  
105 *decreased* rates when the cell's BF is near a formant frequency (Fig. 1F, green arrow) because  
106 the neural inputs have weak low-frequency rate fluctuations (see Fig. 1B). Cells with band-pass  
107 MTFs but with BF between formants are hypothesized to have *increased* rates (Fig. 1F, orange  
108 arrow) because their neural inputs have strong low-frequency fluctuations (see Fig. 1C). In  
109 contrast, cells with band-reject or low-pass MTFs and *minima* near F0 will have *increased* rates

110 when formant frequencies are near BF (Fig. 1G, green arrow), because their neural inputs have  
111 weak low-frequency fluctuations (see Fig. 1B). Band-reject or low-pass cells will have *decreased*  
112 rates (Fig. 1G, orange arrow) when BF is between formants and the inputs have strong  
113 fluctuations (see Fig. 1C).

114       The contrast across frequency in the F0-related neural fluctuations sets up a code for  
115 formants that is translated into rate profiles in the midbrain (Fig. 1F,G). This study used  
116 computational models for AN fibers and IC neurons to explore the robustness of this proposed  
117 code across a wide range of sound levels and in background noise. Examples of recordings from  
118 IC neurons in awake rabbit support the model for cells that have simple band-pass or band-reject  
119 amplitude-modulation tuning.

120

## 121 **Materials and Methods**

122       Modeling: A phenomenological model of AN responses that includes several key  
123 nonlinearities, including rate saturation, adaptation, and synchrony capture (Zilany et al., 2009,  
124 2014) provided the inputs to the models for two types of midbrain neurons (Fig 2A). IC cells  
125 with bandpass (BP) MTFs were simulated using the Same-Frequency Inhibition-Excitation  
126 (SFIE) model (Nelson and Carney, 2004), which explains tuning for amplitude-modulation  
127 frequency by the interaction of excitatory and inhibitory inputs with different dynamics. IC cells  
128 with low-pass, band-reject (LPBR), or high-pass MTFs were simulated using an extension of the  
129 SFIE model; the LPBR model received excitatory input from the brainstem and inhibitory input  
130 from band-pass cells (Fig. 2B). Time-varying input rate functions to each model cell were  
131 convolved with alpha functions representing excitatory or inhibitory post-synaptic responses.  
132 The decay time constants of the alpha functions and delays associated with synaptic responses

133 were varied to produce MTFs tuned to different amplitude-modulation frequencies (Nelson and  
134 Carney, 2004).

135         The parameter values for the cochlear nucleus (CN)/brainstem neurons (Fig. 2) were  
136 fixed for all simulations. These parameters were the time constants of the excitatory and  
137 inhibitory alpha functions,  $\tau_{\text{CNex}}$  and  $\tau_{\text{CNinh}}$ , the delay of the inhibitory alpha function with respect  
138 to the excitatory alpha function,  $D_{\text{CNinh}}$ , and amplitude scalars for the excitatory and inhibitory  
139 alpha functions  $A_{\text{CNex}}$  and  $A_{\text{CNinh}}$  (see Table I for parameter values). These parameters resulted in  
140 “generic” cochlear nucleus or other brainstem neurons that project to the IC. In general, many  
141 types of CN/brainstem neurons have an increased synchrony to the stimulus envelope and weak  
142 rate tuning to the amplitude-modulation rate (e.g. Frisina et al., 1990; Grothe et al., 1997; Joris &  
143 Yin, 1998; Gai & Carney, 2008; reviewed by Joris et al., 2004). The model CN/brainstem cells  
144 do not have significantly greater contrast in the rate vs. BF profiles in response to vowels than do  
145 AN fibers, thus this stage of the model is not an attempt to simulate the CN “chopper” neurons  
146 described by Blackburn & Sachs (1990), which have weaker temporal representations of vowels  
147 than AN fibers or other CN response types but more robust rate vs. BF profiles. The detailed  
148 response properties and connections of different types of CN and other brainstem neurons was  
149 not included in the simulations here.

150         Parameter values for model IC neurons are provided in Table 1. For model bandpass  
151 neurons, there were four parameters: the time constant of the excitatory alpha function,  $\tau_{\text{BPex}}$ , the  
152 time constant of the inhibitory alpha function,  $\tau_{\text{BPinh}}$ , the delay of the inhibition with respect to  
153 the excitation,  $D_{\text{BPinh}}$ , and amplitude scalars for the excitatory and inhibitory inputs,  $A_{\text{BPex}}$  and  
154  $A_{\text{BPinh}}$ . These parameters were based on example model neurons with a range of BMFs in Nelson  
155 & Carney (2004). Model band-reject, low-pass, and high-pass neurons (Fig. 2B) were described

156 by the bandpass model parameters, plus the time constant of the excitatory alpha function,  
157  $\tau_{LPBRex}$ , the time constant of the inhibitory alpha function,  $\tau_{LPBRinh}$ , the delay of the inhibition  
158 with respect to excitation,  $D_{LPBRinh}$ , and amplitude scalars for the excitatory and inhibitory inputs,  
159  $A_{LPBRex}$  and  $A_{LPBRinh}$ .

160 For all models, the synaptic output signal from the auditory-nerve model (which has units  
161 of spikes/sec) was convolved with excitatory and inhibitory post-synaptic potentials for the  
162 CN/brainstem model. These potentials were modeled by alpha functions, each described by a  
163 time constant, and each normalized to have unit area before scaling the amplitudes with the  
164 coefficients described above. The model cell output was computed by subtracting the inhibitory  
165 signal from the excitatory potential and then half-wave rectifying. This model output signal was  
166 then convolved with the appropriate alpha function to provide the input to the next model cell,  
167 and excitatory and inhibitory signals were summed and half-wave rectified to compute the model  
168 IC response.

169 The basic properties of the model responses to the vowel /æ/ (in “had”) are illustrated in  
170 Fig. 2C as a function of time for a population of neurons tuned to a range of best frequencies  
171 (BF, the frequency that elicits the strongest response). As illustrated schematically in Fig. 1, the  
172 model AN fibers (Fig. 2C) tuned near formant frequencies (arrows at right) have strong response  
173 rates with relatively small F0-related fluctuations; and those tuned to intermediate frequencies  
174 have strong fluctuations in rate associated with each pitch-period. A population of model BP  
175 cells with MTF tuned to F0 and a range of BFs is illustrated in Fig. 2D. BP cells with BFs tuned  
176 to formant frequencies have weak responses compared to those tuned to intermediate  
177 frequencies, where the strong F0-related fluctuations elicit strong responses (Fig. 2D, and see  
178 Fig. 1C). In contrast, model LPBR cells (Fig. 2E) with a notch in the MTF near F0 (see Fig. 1E)

179 respond best when BF is tuned near a formant frequency; these cells respond weakly to the  
180 strong F0-related fluctuations on their inputs (see Fig. 1C), and are more effectively driven by  
181 the weaker modulations provided by the auditory periphery at the formant frequencies (See Fig.  
182 1B).

183 Figure 3 illustrates MTFs for three model BP neurons (blue curves) tuned to different  
184 amplitude-modulation rates. Each of these model BP neurons provided an inhibitory input to a  
185 LPBR neuron, yielding the other set of MTFs (red curves). The shapes of these MTFs are  
186 characteristic of MTF types referred to as band-reject, low-pass, and high-pass (e.g. Krishna and  
187 Semple, 2000; Nelson and Carney, 2007). This relatively simple model for modulation frequency  
188 tuning explains several of the MTF types that are encountered in the IC. Other IC cells have  
189 more complex MTFs, consisting of combinations of excitatory and inhibitory regions Krishna  
190 and Semple, 2000). Further extensions of this modeling approach will be required to describe  
191 those MTFs.

192 **Physiological Methods:** All animal procedures were performed in accordance with the  
193 University of Rochester animal care committee's regulations. Recordings were made in the  
194 central nucleus of the IC in female awake rabbits using implanted tetrodes advanced through the  
195 IC with a head-mounted microdrive (Neuralynx 5-Drive, Bozeman, MT). Tetrodes were  
196 constructed by twisting 4 strands of 12 or 18  $\mu\text{m}$  diameter epoxy-coated platinum iridium wire.  
197 Action potentials were sorted off-line based on spike shapes (Schwarz et al., 2012). Single-unit  
198 recordings were identified based on a criterion of  $< 2\%$  for inter-spike intervals less than 1 msec  
199 and, when multiple spikes were recorded and sorted, on values  $< 0.1$  of the summed cluster  
200 separation metric ( $L_{\Sigma}$ , based on a sum of the cluster isolation metric  $L_{\text{ratio}}$ , from Schmitzer-  
201 Torbert et al., 2005, see Schwarz et al., 2012).

202 Acoustic stimuli were created in Matlab and presented using TDT hardware (Tucker-  
203 Davis, Gainesville, FL) and Beyer Dynamic (Farmingdale, NY) headphones through custom  
204 earmolds. Stimuli were calibrated using an Etymotic (Elk Grove Village, IL) ER-7C probe-tube  
205 microphone. Audio frequency tuning was determined using response maps based on responses to  
206 200-ms tones presented every 600 ms with frequencies from 0.2-20 kHz and levels from 10-70  
207 dB SPL, presented in random order. Amplitude-modulation tuning was determined using 100%  
208 modulated wideband noise (30 dB SPL spectrum level) or tone carriers (70 dB SPL) near the  
209 neurons' best frequencies. Vowel stimuli (65 dB SPL) were from the Hillenbrand database  
210 (Hillenbrand et al., 1995). Samples were extracted from the steady-state portion of the vowel,  
211 and a Hanning window was applied to limit the duration to 200 ms. Vowel stimuli were chosen  
212 from the database based on the match of the speaker's average F0 to the neuron's BMF.

213

## 214 **Results**

215 Model Responses:

216 Responses of model IC cells with band-pass and band-reject MTFs to two vowel tokens  
217 (Fig. 4) illustrate the representation of formants in the average rate profiles of model IC  
218 population responses. As schematized in Fig. 1, the profile of average discharge rates for a  
219 population of model IC neurons with BP MTFs (Fig. 4C,D, blue) has minima at BFs near the  
220 vowel formants. In contrast, the rates of band-reject neurons (Fig. 4C,D, red) have peaks at the  
221 formants. The importance of the LPBR model for a robust neural code of vowel formants is  
222 illustrated in Fig. 4D for the vowel /i/, which like many vowels has widely spaced formants. This  
223 response shows that reductions in discharge rate of BP responses (Fig. 4D, blue) are ambiguous,  
224 as they may be due either to reduced fluctuations of AN responses tuned near formants (see Fig.

225 1B) or to reduced spectral energy (Fig. 4D, arrow, 1500 Hz). This ambiguity is resolved by the  
226 LPBR model (Fig. 4D, red), which only responds when *both* sufficient energy and reduced  
227 fluctuations are present on the inputs to the model midbrain cell. The reduced fluctuations result  
228 in disinhibition of the LPBR model by reducing the inhibitory input from the BP neuron. Note  
229 that the model LPBR population rate profile is qualitatively similar to the AN (Fig. 4C, D,  
230 magenta) and CN/brainstem (Fig. 4C,D, cyan) profiles, except that the LPBR population  
231 responses (Fig. 4C, D, red curves) have greater contrast and steeper slopes in the discharge rates  
232 across the population in comparison to the peripheral responses. The LPBR model average rate  
233 profiles differ from peripheral rate profiles in being more robust for vowels in background noise  
234 and across a wide range of sound levels (see below).

235         The midbrain vowel-coding hypothesis is robust across a wide range of SPLs (Fig. 5)  
236 because it is based on the pattern of pitch-related rate fluctuations in AN responses, as opposed  
237 to spectral energy or average rates of AN fibers. Model AN rates, shown in response to the  
238 vowel /æ/ (“had”), saturate at moderate to high sound levels, obscuring the representations of  
239 formant peaks (Fig. 5A). All model responses presented here are based on models for low-  
240 threshold high-spontaneous-rate AN model fibers, which are the majority of AN fibers  
241 (Liberman, 1978). Although responses of medium- or low-spontaneous-rate fibers have  
242 somewhat larger dynamic ranges and higher thresholds, the representation of formant peaks in all  
243 fiber types weakens as sound level increases and the fibers begin to saturate.

244         The representations of F1 and F2 for the vowel /æ/ (“had”) in the average discharge rate  
245 profiles of populations of model IC cells appear as decreased rates for model BP cells tuned near  
246 the formants (vertical blue streaks in Fig. 5B) or increased rates for model LPBR cells (vertical  
247 orange streaks in Fig. 5C). The *contrast* in rates, e.g. the difference between peaks (red) and

248 minima (blue), along the frequency axis varies with SPL. This contrast is strongest for sound  
249 levels near 65 dB SPL (Fig. 5B, 5C); *i.e.*, in the range of conversational speech. The wide  
250 dynamic range of the formant representation is due partly to spike-rate adaptation (Dean et al.,  
251 2005, 2008; Wen et al., 2009, 2012) which increases the overall dynamic range of auditory  
252 neurons, a phenomenon largely explained by the power-law synaptic adaptation included in the  
253 AN model (Zilany & Carney, 2010).

254         The reduction in the contrast of rates in the model responses at high levels is consistent  
255 with the phenomenon of “roll over,” the gradual decrease in speech recognition scores at levels  
256 exceeding 80 dB SPL (Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999). Addition of  
257 smaller percentages of medium- and low-spontaneous-rate AN fibers to the high-spontaneous  
258 model population would slightly increase the model dynamic range, but the representation of  
259 formants would still roll off at the highest levels (not shown). The high-spontaneous-rate AN  
260 models were used as inputs for the IC models shown here to emphasize that the information  
261 required for the wide dynamic range of the proposed coding hypothesis is present even in this  
262 group of AN fibers, which has the smallest dynamic range.

263         The representation of formants in the model midbrain average discharge rate profiles is  
264 also robust in the presence of additive speech-shaped Gaussian noise across a range of signal-to-  
265 noise ratios (SNR) (Fig. 6). Figure 6A shows model AN fibers in response to the vowel /æ/  
266 (“had”); as SNR decreases, the representation of the formants in the AN discharge rates  
267 deteriorates, especially in the F2 frequency region. Formant representation is much more robust  
268 in the response profiles of midbrain neurons (Fig. 6B, 6C). The dips in the response profile of the  
269 model BP cells (Fig. 6B) and the peaks in the LPBR profile (Fig. 6C) deteriorate at  
270 approximately the speech reception threshold (SRT), where human listeners have difficulty

271 understanding with speech in noise ( $\sim -5$  dB SNR, Festen & Plomp, 1990).

272

273 Physiological Responses:

274         The vowel-coding model was tested with recordings from IC neurons in awake rabbit to  
275 twelve contrastive English vowels from one human male speaker with an average F0 of 128 Hz  
276 (Hillenbrand et al., 1995). Responses of 75 neurons with BFs below 4500 Hz that responded to  
277 65 dB SPL vowel stimuli were compared to model predictions; a subset of these neurons were  
278 also studied at multiple SPLs and SNRs.

279         Figure 7 illustrates responses of two neurons, one with a BF of 1100 Hz and a band-pass  
280 MTF (Fig. 7A) and the other with a BF of 2000 Hz and a band-reject MTF (Fig. 7B). Figure 7C  
281 and D show average discharge rates for these two cells in response to 9 English vowels (black  
282 line), along with predictions provided by the BP SFIE (blue line) and LPBR (red line) models  
283 (see Figs. 1, 4). For comparison, predictions based on the energy through a gammatone filter  
284 centered at BF are also shown (green line). The Pearson product moment correlation coefficient  
285 between actual rates and each of the predictions is also shown.

286         Discharge rates of the BP cell were not explained by the stimulus energy near the  
287 neuron's BF. For example, for the cell in Fig. 7A,C, the energy near BF is greatest for the vowels  
288 /b/ (in "father") and /ɔ:/ (in "bawd"), yet the neuron's discharge rates are low in response to these  
289 vowels. In contrast, the neuron responds strongly to /i/ (in "heed") and /ɪ/ (in "hid"), which have  
290 relatively low energy near this neuron's BF (Fig. 7C). The BP SFIE model, however, explains  
291 these counterintuitive responses of the BP IC neurons to vowels (cf. blue and black lines in Fig.  
292 7C). The BP neuron's responses *decreased* when formant frequencies encroached upon the  
293 neuron's best frequency (1300 Hz), as predicted by the SFIE model (Fig. 1F, 4C, blue), because

294 of the *reduced rate fluctuations* in the those frequency channels. Synchrony capture and  
295 saturation of AN fibers tuned near the formant peaks result in reduced rate fluctuations in the  
296 responses of those frequency channels. Knowledge of the neuron's BF, MTF type, and BMF  
297 allowed predictions of the BP cell's vowel responses by the SFIE BP model.

298         The band-reject neuron's (Fig. 7D, black) responses increased when formant frequencies  
299 were near the neuron's BF (2000 Hz), as predicted by the LPBR model (Fig 7D, red). Although  
300 the band-reject neuron's responses were positively correlated to energy near BF (Fig. 7D, green),  
301 the LPBR model responses also reflected trends in the band-reject neuron's responses that were  
302 not explained solely by the stimulus energy in the critical band centered at BF.

303         An important property of the proposed model for vowel coding is its resilience across  
304 SPL (Fig. 5) and SNR (Fig. 6). Some cells in the IC have discharge rate profiles that are  
305 similarly robust across a wide range of stimulus parameters. An example is shown in Fig. 8. This  
306 neuron had a band-reject MTF (Fig. 8A) and its discharge rates in response to the set of nine  
307 vowels presented at 65 dB SPL were well predicted by the LPBR model and by the energy  
308 model (Fig. 8B). The large differences in rate across the set of vowels for this neuron facilitate  
309 comparisons of the rate profile across a range of SPLs (Fig. 8C) and SNRs (Fig. 8E). As SNR  
310 decreases, the rate profile approaches the response to 65 dB noise alone (Fig. 8E, blue), whereas  
311 at high SNRs the profile approaches the response to speech in quiet (Fig. 8E, black). For  
312 comparison, responses of a high-spontaneous-rate model AN fiber with the same BF (1100 Hz)  
313 are shown for the same range of SPLs (Fig. 8D) and SNRs (Fig. 8E). The AN rates across this set  
314 of vowels gradually saturate over this range of sound levels (Fig. 8D). All of the AN responses  
315 for stimuli that included the added speech-shaped noise were saturated for the SNRs studied  
316 (Fig. 8E).

317           The physiological results above demonstrate examples of IC responses with features that  
318 are consistent with the model. Of 75 neurons that responded to 65 dB vowel stimuli with F0 in  
319 the 100-130 Hz range, 62 neurons (83%) had average rates in response to a set of 12 vowels that  
320 were significantly correlated (i.e.  $r \geq 0.57$ , 2df) by at least one of the three models (BP, LPBR,  
321 or Energy). Of these, 11% were best predicted by the BP model, and 42% were best predicted by  
322 the LPBR model. Note that many neurons in the IC have more complex MTFs than the simple  
323 band-pass and band-reject examples shown above. In particular, MTFs that combine excitatory  
324 and inhibitory regions at different modulation frequencies are common (Krishna and Semple,  
325 2000), and further extension of the model is required to describe those neurons' responses to  
326 vowels. Figure 9 illustrates diverse MTFs, vowel responses, and correlations to model  
327 predictions for five additional IC neurons. These complex MTF shapes illustrate the challenge of  
328 classifying neurons as simply "bandpass" or "band-reject." Each of these neurons has rates that  
329 are enhanced and/or suppressed with respect to the response to the lowest modulation frequency  
330 tested. Kim et al. (2015) propose categorization of MTFs as band-enhanced or band-suppressed,  
331 based on comparisons to the response to an unmodulated stimulus. The examples in Fig. 9 have  
332 responses that are sometimes better predicted by the BP model (Fig. 9A,D), and sometimes by  
333 the LPBR model (Fig. 9B,C,E). However, it should be noted that in some cases (e.g. Fig. 9A),  
334 the correlation between model and neural responses is strongly influenced by the responses to  
335 one or two vowels. The correlations in Fig. 9 also illustrate that although the LPBR and energy  
336 model responses are often highly correlated (e.g. Figs. 7, 8, 9B,C,E), this is not always the case  
337 (e.g. Fig. 9A,D). In general, for the examples in Fig. 9 the BP model provides better predictions  
338 of responses for neurons that have peaks in the MTF near the F0 of the stimulus, and the LPBR  
339 provides better predictions when there is a dip in the MTF near F0. Thus, it is reasonable to

340 hypothesize that quantifying the neural fluctuations established in the periphery near a neuron's  
341 BF, and then applying the features of the MTF at modulation frequencies relevant to the  
342 stimulus, will explain the vowel responses for cells with complex MTFs. This strategy provides a  
343 novel and general framework for understanding how complex sounds with strong fluctuations,  
344 such as voiced speech, are encoded at the level of the midbrain.

345

#### 346 **Discussion**

347       Previous studies of the neural representation of vowels have largely focused on coding of  
348 spectral energy by the AN (Delgutte & Kiang, 1984; Liljencrants, 1972; Lindblom, 2009;  
349 Young, 2008). Codes based on AN average discharge rates and/or temporal fine structure are  
350 problematic because of limited dynamic range and background noise. Many AN models,  
351 especially those used in the phonetics literature, are based on linear filter banks (e.g., Ghosh et  
352 al., 2011; Lindblom, 1990; Diehl et al, 2004). The model presented here, in contrast, includes the  
353 nonlinear attributes of AN responses, including level-dependent tuning bandwidth, synchrony  
354 capture, and saturation, all of which influence the neural fluctuations in response to speech  
355 sounds that ultimately project to the IC. The coding hypothesis here focuses on the F0-related  
356 fluctuations in the AN responses and how they vary across the population. These fluctuations are  
357 important because IC neurons are tuned to both audio and modulation frequencies. This tuning  
358 provides sensitivity to the contrast in low-frequency, pitch-related, fluctuations across a  
359 population of neurons with different best frequencies (cf. Fig. 1B and 1C).

360       IC responses tuned to a range of modulation frequencies encode vowel formant  
361 frequencies across a range of pitches, capturing an essential aspect of speech (Diehl, 2008).

362 This model framework provides a context for understanding several general questions related to  
363 vowel systems, which exhibit universal properties that generalize across languages. For example,  
364 formant locations appear to disperse optimally within an F1-F2 space, referred to as dispersion  
365 theory (Liljencrants & Lindblom, 1972; Lindblom, 1986; Diehl & Lindblom, 2004; Schwartz et  
366 al., 1997). This dispersion results in consistencies among linguistic vowel systems in the  
367 appearance of vowel contrasts as vowel systems increase in size. Our model for neural  
368 representations of vowels thus provides a new tool for understanding how the role of the auditory  
369 system in shaping vowel systems.

370 This model also provides a framework to study the relative spacing of formant peaks, F1,  
371 F2, and F3, which define single vowels. The neural resolution for coding separate formants, or  
372 for merging them, depends upon both the separation of the formant peaks and the widths of the  
373 formant bands. Limits in neural resolution along the frequency axis for coding single formant  
374 peaks would determine when nearby formants merge perceptually (Chistovich & Lublinskaya,  
375 1979). This concept underlies focalization-dispersion theory (Schwartz et al, 1997; Krystal-  
376 Becker, 2010). In the neural model, the width of a single formant's representation along the  
377 frequency axis (e.g., Figs. 1F, 1G, 2, and 6) depends upon the width of the modulation transfer  
378 functions for these neurons (Figs. 1D, 1E, and 4). Future studies to test the hypothesis presented  
379 here should include synthetic vowel sounds, in which the key parameters can be systematically  
380 manipulated with respect to the frequency and modulation tuning of a given neuron, as well as  
381 higher-level features such as formant spacing. These stimuli would also provide allow stronger  
382 statistical tests of the correlations between model and neural responses than was possible with  
383 the small set of spoken vowels used in this study.

384           Recent results in the cortex suggest that phonetic features are encoded in the responses of  
385 the superior temporal gyrus (Mesgarani et al., 2008; Pasley et al., 2012) but the problem of how  
386 neural maps at cortical levels are created from the acoustic stimulus remains. The results  
387 presented here suggest a framework for sub-cortical neural coding of phonetic features based on  
388 patterns of F0-related neural fluctuations. These patterns are established in the auditory periphery  
389 by the nonlinear response properties of inner-ear mechanics and sensory transduction. Contrasts  
390 in these patterns are then enhanced by the sensitivity of midbrain neurons to fluctuation  
391 frequencies. The potential also exists for amplification of these contrasts in the thalamus and  
392 cortex by interactions between inputs from band-pass and band-reject midbrain responses.  
393 Responses of midbrain cells with complex MTFs, characterized by a combination of inhibition  
394 and excitation (e.g. Krishna and Semple, 2000), could serve as effective “edge-detectors,” further  
395 enhancing the contrasts in rate fluctuations across the neural population. In general, the  
396 combination of spectral frequency tuning and modulation frequency processing in the midbrain  
397 provides a substrate for parsing complex sounds into the features that are required for higher-  
398 level phonetic representations.

399           The stimuli modeled and tested in this study were limited to vowels, and the voiced  
400 structure of these sounds has a strong influence on the responses. It is interesting to consider how  
401 the properties of these neurons would influence responses to other types of speech sounds.  
402 Unvoiced vowels exist in whispered speech, and in vo-coded speech, such as that used in  
403 cochlear implant simulations (Shannon et al., 1995), as well as conditioned alternates of vowels  
404 in several languages (Ladefoged & Maddieson, 1996). Unvoiced or de-voiced vowels have  
405 reduced intelligibility compared to normal vowels. The model presented here would respond  
406 mainly to the energy profile in unvoiced vowels, such that formants would be coded by increased

407 rates for neurons tuned near formants. These energy-related responses would be correlated with  
408 the representation of voiced vowels in the LPBR model, though with reduced contrast in rate as a  
409 function of frequency. Consonants represent another diverse and important set of speech sounds,  
410 sets of which (obstruents) are commonly voiceless (p, t, k, ch), and sometimes characterized by a  
411 noisy source (e.g. fricatives) (Stevens, 1998; Ladefoged, 2006). Similar to vowels, the  
412 consonants set up a pattern of neural fluctuations in the peripheral response that will ultimately  
413 drive the responses at the level of the midbrain. Future studies will extend the general approach  
414 presented here to include a larger set of speech sounds. An interesting question is how midbrain  
415 neurons with different MTFs will represent the slopes and peaks in consonant spectra, which  
416 result in non-periodic but potentially strong fluctuations in peripheral responses.

417         The vowel-coding hypothesis presented here has implications for several applications  
418 related to speech processing. Accurate formant identification in the presence of substantial  
419 background noise is critical for automatic speech-recognition systems, yet is difficult to achieve.  
420 The hypothesis also provides a new framework for speech-enhancement algorithms for listeners  
421 with and without hearing loss. The code is substantially affected by common aspects of hearing  
422 loss, such as broadened frequency tuning in the inner ear, which distorts the representation of  
423 both the spectrum and amplitude modulations. The proposed code is also affected by changes in  
424 synchrony capture that would accompany hearing loss. Loss of synchrony capture has a profound  
425 effect on the nature of the neural fluctuations in peripheral responses, and thus on the responses  
426 of central neurons that are driven by these fluctuations. The hypothesis thus informs  
427 development of algorithms for new hearing aids and cochlear-implant speech processors that  
428 encode information in temporal envelopes.

429

430 **References**

431 Becker-Kristal R (2010) *Acoustic typology of vowel inventories and Dispersion Theory: Insights*  
432 *from a large cross-linguistic corpus*. PhD. Dissertation, UCLA.

433

434 Chistovich LA, Lublinskaya VV (1979) The "center of gravity" effect in vowel spectra and  
435 critical distance between the formants: psychoacoustical study of the perception of vowel-like  
436 stimuli. *Hear Res* **1**:185-195.

437

438 Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to  
439 stimulus statistics. *Nat Neurosci* **8**:1684-1689.

440

441 Dean I, Robinson BL, Harper NS, McAlpine D (2008) Rapid neural adaptation to sound level  
442 statistics. *J Neurosci* **28**:6430-6438.

443

444 Delgutte B, Hammond BM, Cariani PA (1998) Neural coding of the temporal envelope of  
445 speech: relation to modulation transfer functions. In: *Psychophysical and physiological advances*  
446 *in hearing* (Palmer AR, Summerfield AQ, Meddis R, eds.), pp. 595–603. London, UK: Whurr  
447 Publishers.

448

449 Delgutte B, Kiang NY (1984) Speech coding in the auditory nerve: I. Vowel-like sounds. *J*  
450 *Acoust Soc Am* **75**:866-878.

451

- 452 Deng L, Geisler CD (1987) Responses of auditory-nerve fibers to nasal consonant-vowel  
453 syllables. *J Acoust Soc Am* **82**:1977-1988.  
454
- 455 Diehl RL (2008) Acoustic and auditory phonetics: the adaptive design of speech sound systems.  
456 *Philos Trans R Soc Lond B Biol Sci* **363**:965-978.  
457
- 458 Diehl RL, Lindblom B, Creeger CP (2003) Increasing realism of auditory representations yields  
459 further insights into vowel phonetics. In: *Proc. Int. Congr. Phon. Sci., 15th, Barcelona*, vol. 2,  
460 pp. 1381-1384. Adelaide: Causal Publications.  
461
- 462 Diehl R, Lindblom B (2004) Explaining the structure of feature and phoneme inventories: The  
463 role of auditory distinctiveness. In: *Speech Processing in the Auditory System (Greenberg S,*  
464 *Ainsworth WA, Popper AN, Fay RR, eds.)*, pp. 101-162; Springer.  
465
- 466 Fant G (1960) *Acoustic Theory of Speech Production*. Hague, The Netherlands: Mouton.  
467
- 468 Festen JM, Plomp R (1990) Effects of fluctuating noise and interfering speech on the speech-  
469 reception threshold for impaired and normal hearing. *J Acoust Soc Am* **88**:1725-1736.  
470
- 471 Frisina RD, Smith, RL, Chamberlain SC (1990) Encoding of amplitude modulation in the gerbil  
472 cochlear nucleus: I. A hierarchy of enhancement. *Hearing research*, **44**: 99-122.  
473

- 474 Gai Y, Carney LH (2008) Influence of inhibitory inputs on rate and timing of responses in the  
475 anteroventral cochlear nucleus. *Journal of neurophysiology*, **99**:1077-1095.  
476
- 477 Ghosh PK, Goldstein LM, Narayanan SS (2011) Processing speech signal using auditory-like  
478 filterbank provides least uncertainty about articulatory gestures. *J Acoust Soc Am* **129**:4014-  
479 4022.  
480
- 481 Grothe B, Park TJ, Schuller G (1997) Medial superior olive in the free-tailed bat: response to  
482 pure tones and amplitude-modulated tones. *Journal of neurophysiology*, **77**:1553-1565.  
483
- 484 Hillenbrand J, Getty LA, Clark MJ, Wheeler K (1995) Acoustic characteristics of American  
485 English vowels. *J Acoust Soc Am* **97**:3099-3111.  
486
- 487 Joris PX, Schreiner CE, Rees A (2004) Neural Processing of Amplitude-Modulated Sounds.  
488 *Physiol Rev* **84**:541-577.  
489
- 490 Joris PX, Yin TC (1998) Envelope coding in the lateral superior olive. III. Comparison with  
491 afferent pathways. *Journal of Neurophysiology*, **79**:253-269.  
492
- 493 Kim DO, Zahorik P, Carney LH, Bishop BB, Kuwada S (2015) Auditory Distance Coding in  
494 Rabbit Midbrain Neurons and Human Perception: Monaural Amplitude Modulation Depth as a  
495 Cue. *The Journal of Neuroscience*, **35**: 5360-5372.  
496

- 497 Krishna BS, Semple MN (2000) Auditory Temporal Processing: Responses to Sinusoidally  
498 Amplitude-Modulated Tones in the Inferior Colliculus. *J Neurophysiol* **84**:255-273.  
499
- 500 Ladefoged P (2006) *A Course in Phonetics*. Thompson-Wadsworth.  
501
- 502 Ladefoged P, Maddieson I (1996) *Sounds of the World's Languages*. Wiley-Blackwell.  
503
- 504 Langner G (1992) Periodicity coding in the auditory system. *Hear Res* **60**:115-142.  
505
- 506 Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. *J Acoust*  
507 *Soc Am.* **63**:442-455.  
508
- 509 Liljencrants J, Lindblom B (1972) Numerical simulation of vowel quality systems: the role of  
510 perceptual contrast. *Language* **48**: 839-862.  
511
- 512 Lindblom B (1990) Explaining phonetic variation: a sketch of the H & H theory. In: *Speech*  
513 *production and speech modeling (Hardcastle WJ, Marchal A, eds.)*, pp. 403–439. Dordrecht, The  
514 Netherlands: Kluwer.  
515
- 516 Lindblom B (1986) Phonetic universals in vowel systems. In: *Experimental Phonology*. (Ohala J,  
517 Jaeger J, eds.) Orlando: Academic Press, pp. 13-44.  
518

- 519 Lindblom B, Diehl R, Creeger C (2009) Do 'Dominant Frequencies' explain the listener's  
520 response to formant and spectrum shape variations? *Speech Comm* **51**:622-629.  
521
- 522 Mesgarani N, David SV, Fritz JB, Shamma SA (2008) Phoneme representation and classification  
523 in primary auditory cortex. *J Acoust Soc Am* **123**:899-909.  
524
- 525 Miller RL, Schilling JR, Franck KR, Young ED (1997) Effects of acoustic trauma on the  
526 representation of the vowel [bold /epsilon/] in cat auditory nerve fibers. *J Acoust Soc Am* **101**:  
527 3602-3616.  
528
- 529 Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural  
530 responses to amplitude-modulated tones. *J Acoust Soc Am* **116**:2173-2186.  
531
- 532 Nelson PC, Carney LH (2007) Neural Rate and Timing Cues for Detection and Discrimination of  
533 Amplitude-Modulated Tones in the Awake Rabbit Inferior Colliculus. *J Neurophysiol* **97**:522-  
534 539.  
535
- 536 Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF  
537 (2012) Reconstructing speech from human auditory cortex. *PLoS biology* **10**:e1001251.  
538
- 539 Sachs MB, Abbas PJ (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-  
540 burst stimuli. *J Acoust Soc Am* **56**:1835-1847.  
541

- 542 Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve:  
543 Representation in terms of discharge rate. *J Acoust Soc Am* **66**:470-479.  
544
- 545 Schilling JR, Miller RL, Sachs MB, Young ED (1998) Frequency-shaped amplification changes  
546 the neural representation of speech with noise-induced hearing loss. *Hear Res* **117**:57-70.  
547
- 548 Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD (2005) Quantitative measures of  
549 cluster quality for use in extracellular recordings. *Neuroscience*, **131**:1-11.  
550
- 551 Schwartz JL, Boë LJ, Vallée N, Abry C (1997) The Dispersion-Focalization Theory of vowel  
552 systems. *J Phonetics* **25**:255-286.  
553
- 554 Schwarz DM, Zilany MS, Skevington M, Huang NJ, Flynn BC, Carney LH (2012) Semi-  
555 supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. *J*  
556 *Neurosci Methods* **206**:120-131.  
557
- 558 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with  
559 primarily temporal cues. *Science* **270**: 303-304.  
560
- 561 Stevens K (1998) *Acoustic Phonetics*. MIT Press.  
562
- 563 Studebaker GA, Sherbecoe RL, McDaniel DM, Gwaltney CA (1999) Monosyllabic word  
564 recognition at higher-than-normal speech and noise levels. *J Acoust Soc Am* **105**:2431-2444.

565

566 Wen B, Wang GI, Dean I, Delgutte B (2009) Dynamic range adaptation to sound level statistics  
567 in the auditory nerve. *J Neurosci* **29**:13797-13808.

568

569 Wen B, Wang GI, Dean I, Delgutte B (2012) Time course of dynamic range adaptation in the  
570 auditory nerve. *J Neurophysiol* **108**:69-82.

571

572 Yates GK (1990) Basilar membrane nonlinearity and its influence on auditory nerve rate-  
573 intensity functions. *Hear Res* **50**:145-162.

574

575 Yates GK, Winter IM, Robertson D (1990) Basilar membrane nonlinearity determines auditory  
576 nerve rate-intensity functions and cochlear dynamic range. *Hear Res* **45**:203-219.

577

578 Young ED (2008) Neural representation of spectral and temporal information in speech. *Philos*  
579 *Trans R Soc Lond B Biol Sci* **363**:923-945.

580

581 Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of  
582 the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc Am* **66**:1381-1403.

583

584 Zilany MS, Bruce IC, Nelson PC, Carney LH (2009) A phenomenological model of the synapse  
585 between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J*  
586 *Acoust Soc Am* **126**:2390-2412.

587

588 Zilany MS, Carney LH (2010) Power-law dynamics in an auditory-nerve model can account for  
589 neural adaptation to sound-level statistics. *J Neurosci* **30**:10380-10390.

590

591 Zilany, M.S.A., Bruce, I.C., and L. H. Carney (2014) Updated parameters and expanded  
592 simulation options for a model of the auditory periphery. *J Acoust Soc Am* 135:283-286.

593

594

595

596

597

598

599

600

601 Legends

602 Figure 1. Schematic illustration of vowel-coding hypothesis. The left-hand column labels the key  
603 stages in the coding scheme. A) Vowel spectrum consisting of harmonics of F0, shaped by the  
604 spectral envelope. B) Responses of AN fibers tuned *near* formants have relatively small pitch-  
605 related rate fluctuations. These responses are dominated by a single harmonic in the stimulus,  
606 referred to as synchrony capture. C) Fibers tuned *between* formants have strong rate fluctuations  
607 at F0 (Delgutte and Kiang, 1984). D) Example of a band-pass MTF from rabbit IC with a BMF  
608 near F0 for a typical male human speaker. E) Example band-reject MTF with a notch near a  
609 typical F0. F) Band-pass midbrain neurons have *reduced* rates in frequency channels with weak  
610 fluctuations (green arrow) and increased rates in channels with strong fluctuations (see C, orange  
611 arrow); thus *dips* in the rate profile of band-pass neurons encode F1 and F2. G) The profile of  
612 rates across a population of band-reject neurons has *peaks* at F1 and F2, because band-reject  
613 neurons respond more strongly to stimuli that result in reduced neural fluctuations in their inputs  
614 (see B, green arrow).

615

616 Figure 2. Models for modulation tuning in IC cells. A) Time waveform of the vowel /æ/ (“had”).  
617 B) The SFIE model (Nelson and Carney, 2004) for midbrain cells with bandpass (BP) MTFs  
618 (blue cell). An extension of the SFIE model is illustrated by the red cell, which is excited by  
619 ascending inputs and inhibited by the bandpass SFIE cell. This model cell simulates the  
620 relatively common low-pass or band-reject MTFs (see Fig. 3), and is referred to as the LPBR  
621 model. C) Model AN population response (Zilany et al., 2009, 2014). D) Population response of  
622 the BP IC model; BP neurons with BF’s near F1 and F2 (arrows at right) have *decreased*

623 responses (see Fig. 1F). E) The LPBR model has *peaks* in the population rate profile near F1 and  
624 F2 (see Fig. 1G).

625

626 Figure 3. Three band-pass MTFs (blue, as in Fig. 2) with A) mid-, B) high- and C) low-  
627 frequency BMFs. MTFs for three model cells (red, as in Fig. 2) that are inhibited by the band-  
628 pass cells explain three other MTF types in the IC: the more common A) band-reject and B) low-  
629 pass MTFs, as well as the less common C) high-pass MTF. Model parameters are in Table I.

630

631 Figure 4. Model IC responses to vowel tokens. Spectra of actual vowels: A) /æ/, B) /i/. C, D)  
632 Rate profile of model cells with BP MTFs tuned to F0 (blue) has dips at formants (dots). Rate  
633 profile of LPBR model cells with minima in MTF near F0 (red) has peaks near formants (dots).  
634 AN (magenta) and CN/brainstem (cyan) rate profiles. D) Rate profile for BP cells has ambiguous  
635 dip (arrow) for vowels with a broad spectral valley. LPBR cells (red) have relatively low rates  
636 where spectral energy is low, due to reduced excitatory inputs. Vowel levels were 65 dB SPL.  
637 Model parameters are the same as Fig. 3B.

638

639 Figure 5. Population rate profiles for model AN (A), BP (B) and LPBR (C) cells in response to  
640 the vowel /æ/ (“had”) for a range of sound levels. Vertical dotted lines indicate the first two  
641 formant frequencies. A) Peaks (red) in AN rates that code formants at low SPLs saturate as level  
642 increases. B) Dips (blue) in the rate profile for F1 and F2 extend from ~30 to 95 dB SPL and are  
643 strongest at conversational speech levels, 50-70 dB SPL. C) LPBR model cells have peaks (red)

644 in the rate profile at the formant frequencies; these peaks are most robust at conversational  
645 speech levels. Model parameters are the same as Fig. 3B.

646

647 Figure 6. Population rate profiles for model AN (A), BP (B), and LPBR (C) cells in response to  
648 the vowel /æ/ (“had”) for a range of signal-to-noise ratios (SNRs). Vowel levels were fixed at 65  
649 dB SPL; noise level increases towards bottom of plots. A) Saturation of AN rates by the added  
650 noise obscures representations of formant frequencies, especially in the F2 region. B) Dips in the  
651 average discharge rate profile that indicate the first two formants in the BP population response  
652 deteriorate gradually as SNR decreases (towards bottom of plot). C) Peaks in the rate profile *vs.*  
653 SNR for model LPBR cells also deteriorate as SNR decreases. Arrow and horizontal dashed lines  
654 indicate approximate speech reception threshold (SRT) for normal-hearing listeners (Festen &  
655 Plomp, 1990). Model parameters are the same as Fig. 3B.

656

657 Figure 7. Examples of two IC neurons in awake rabbit with A) band-pass MTF (BF = 1300 Hz,  
658 BMF = 130 Hz) and B) band-reject MTF (BF=2000 Hz, MTF notch at 150 Hz). C) Black:  
659 Average rate of the band-pass neuron in response to 9 vowels with F0=148 Hz (Hillenbrand et  
660 al., 1995), 65 dB SPL. Blue: Responses of the band-pass SFIE model. Red: LPBR model  
661 responses. Green: Energy at the output of a 4<sup>th</sup>-order gammatone filter at the cell’s BF. Mean and  
662 standard deviation of model responses were matched to neural responses. Lines connect the  
663 symbols to emphasize patterns in the responses across this set of vowels. D) Average rate of the  
664 band-reject neuron (black) to vowels with F0=95 Hz presented at 55 dB SPL, with LPBR model  
665 predictions (red), energy (green), and for comparison, the SFIE model response (blue). Model  
666 parameters were the same as Fig. 3B.

667

668 Figure 8. Example of a neuron (BF = 1100 Hz) with a band-reject MTF (A) for which average  
669 discharge rates in response to 65 dB SPL vowels were best predicted by the LPBR model or the  
670 energy model (B). This neuron's patterns of average discharge rate across the set of vowels were  
671 consistent across a range of SPLs (C) and SNRs (E). Vowel responses for a model AN fiber with  
672 BF=1100 Hz is shown for the same range of SPLs (D) and SNRs (F). Vowel F0 for all datasets  
673 was 95 Hz. IC Model parameters were the same as Fig. 3B.

674 Figure 9. Example of 5 neurons with diverse MTFs (left panel) and predictions of responses to 9  
675 English vowels (right panel) at 65 dB SPL, with correlations to the model predictions in the  
676 legends. BFs were A) 3900, B) 2700, C) 1900, D) 4020, and E) 1485 Hz. Model parameters were  
677 the same as Fig. 3B.

678 Table I. Model Parameters. A single set of parameters was specified for the CN/brainstem level  
 679 of the model. Three sets of parameters were used for the IC models, illustrated in Fig. 3, which  
 680 had BMFs of A) 45 Hz, B) 125 Hz, and C) 16 Hz. The model parameters for other figures were  
 681 the same as for Fig. 3B, which had a BMF near F0 for the most of the vowels used as stimuli.

682

CN/Brainstem (all simulations)			
$\tau_{ex}$ (ms)	0.5		
$\tau_{inh}$ (ms)	2		
D (ms)	1		
$A_{ex}$	1.5		
$A_{inh}$	0.9		
IC model parameters			
	Fig.3A	Fig. 3B	Fig. 3C
Band-Pass Model			
$\tau_{BPex}$ (ms)	2	0.7	5
$\tau_{BPinh}$ (ms)	6	0.7	10
$D_{BP}$ (ms)	2	1.4	2
$A_{BPex}$	2	3	6
$A_{BPinh}$	2.2	4.2	6.6
Low-Pass/Band-Reject Model			
$\tau_{LPBReX}$ (ms)	2	0.7	5
$\tau_{LPBRinh}$ (ms)	5	5	5
$D_{LPBR}$ (ms)	0.7	0.7	0.7
$A_{LPBReX}$	0.6	1	0.6
$A_{LPBRinh}$	2	2	2

683  
 684  
 685

















