

# Speech Coding in the Midbrain: Effects of Sensorineural Hearing Loss

Laurel H. Carney, Duck O. Kim and Shigeyuki Kuwada

**Abstract** In response to voiced speech sounds, auditory-nerve (AN) fibres phase-lock to harmonics near best frequency (BF) and to the fundamental frequency (F0) of voiced sounds. Due to nonlinearities in the healthy ear, phase-locking in each frequency channel is dominated either by a single harmonic, for channels tuned near formants, or by F0, for channels between formants. The alternating dominance of these factors sets up a robust pattern of F0-synchronized rate across best frequency (BF). This profile of a temporally coded measure is transformed into a mean rate profile in the midbrain (inferior colliculus, IC), where neurons are sensitive to low-frequency fluctuations. In the impaired ear, the F0-synchronized rate profile is affected by several factors: Reduced synchrony capture decreases the dominance of a single harmonic near BF on the response. Elevated thresholds also reduce the effect of rate saturation, resulting in increased F0-synchrony. Wider peripheral tuning results in a wider-band envelope with reduced F0 amplitude. In general, sensorineural hearing loss reduces the *contrast* in AN F0-synchronized rates across BF. Computational models for AN and IC neurons illustrate how hearing loss would affect the F0-synchronized rate profiles set up in response to voiced speech sounds.

**Keywords** Vowels · Modulation transfer function · Neural coding

## 1 Introduction

The impact of sensorineural hearing loss (SNHL) on speech communication is perhaps the most significant problem faced by listeners with hearing loss. A puzzle regarding the effects of SNHL is that relatively small amounts of loss have a surprisingly large impact on listeners, especially in noisy acoustic environments. Here

---

L. H. Carney (✉)

Departments of Biomedical Engineering, Neurobiology & Anatomy,  
Electrical & Computer Engineering, University of Rochester, Rochester, NY, USA  
e-mail: Laurel.Carney@Rochester.edu

D. O. Kim · S. Kuwada

Department of Neuroscience, University of Connecticut Health Center, Farmington,  
CT, USA

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894,  
DOI 10.1007/978-3-319-25474-6\_45

we focus on the effects of SNHL on the representation of voiced speech sounds in the auditory periphery and in the midbrain (inferior colliculus, IC). Responses of models for auditory-nerve (AN) fibres and IC neurons provide insight into how changes in the temporal structure of AN responses impact representations at the level of the IC. Neurons in the IC are sensitive to low-frequency rate fluctuations, including rates synchronized to envelopes of amplitude-modulated (AM) stimuli.

Vowels are an essential component of all languages, and play an important role in carrying information, especially in running speech (e.g. Kewley-Port et al. 2007). Vowel sounds are characterized by voicing, which results in a spectrum with a fundamental frequency ( $F_0$ ), related to voice pitch, and harmonics of  $F_0$ . The spectrum is shaped by vocal tract resonances, or formants. The locations of the lowest two frequency formants ( $F_1$  and  $F_2$ ), distinguish different vowel sounds (Fant 1960). Neural coding of these formants is essential for understanding how the brain processes speech.

Neural studies of vowel coding have focused on rate-place and fine-structure temporal coding of formants in AN responses (e.g. Sachs and Young 1979; Young and Sachs 1979; Delgutte and Kiang 1984; reviewed by Young 2008). Consideration of midbrain responses to vowels shifts the focus to the peripheral representation of low-frequency fluctuations ( $\approx 250$  Hz) associated with  $F_0$ . IC cells are particularly sensitive to amplitude modulation in the frequency range of  $F_0$  (Langner 1992). In the healthy ear, nonlinear AN response properties result in systematic patterns of neural fluctuations synchronized to  $F_0$  or higher harmonics near AN characteristic frequencies. In particular, synchrony capture results in the dominance of a single harmonic near a spectral peak on the fine-structure of AN responses to voiced sounds (Delgutte and Kiang 1984; Deng and Geisler 1987). After trauma-induced SNHL, synchrony capture is reduced, and AN fibres respond to multiple harmonics (Miller et al. 1997). These studies focused how synchrony capture effects fine-structure coding; however, this mechanism also sets up a pattern of low-frequency fluctuations across the AN population. In the healthy ear, the fibres that are “captured” by a single harmonic and have responses with relatively flat envelopes. In contrast, fibres tuned to frequencies between formants respond to multiple harmonics and strong rate fluctuations at  $F_0$ . The amplitude of  $F_0$ -related neural fluctuations provides a robust code for the formant frequencies. In the ear with SNHL, decreased synchrony capture results in AN fibres that respond to multiple harmonics, thus  $F_0$ -related fluctuations prevail across all driven frequency channels. The profile of  $F_0$ -related fluctuations that codes the formants, and that ultimately drives modulation-sensitive midbrain neurons, is diminished by SNHL.

In this study, computational models illustrate vowel responses of AN fibres with and without SNHL. Next, models for three types of modulation sensitivity in the IC illustrate the effect of SNHL on population responses in the midbrain.

## 2 Methods

The Zilany et al. (2014) model for AN responses has a provision for including impairment due to outer and inner hair cell dysfunction (Zilany and Bruce 2007), which reduces synchrony capture in a manner comparable to that observed in acoustically traumatized ears (Miller et al. 1997). Here SNHL was simulated by setting AN model parameters  $C_{OHC}$  and  $C_{IHC}$  to 0.3; these values simulate mild hearing loss, with threshold elevations ranging from about 15 dB at 500 Hz to 30 dB at 2000 Hz and higher.

Phenomenological models were used to simulate the three types of rate modulation transfer functions (MTFs) that comprise the bulk of IC neurons (Kim et al. 2015a): band-enhanced, band-suppressed, and hybrid (Fig. 1). The first two MTF types have discharge rates to AM stimuli that are enhanced or suppressed relative to responses to unmodulated stimuli (Kim et al. 2015a, 2015b). Hybrid MTFs have discharge rates that are enhanced at some modulation frequencies and suppressed at others (Krishna and Semple 2000; Kim et al. 2015a).

The band-enhanced model (Fig. 1, BE) was Nelson and Carney's (2004) same-frequency inhibitory-excitatory (SFIE) model. The key mechanism in this model is a dynamic interaction between relatively short-duration excitatory and relatively long-duration, delayed inhibitory synaptic potentials. The durations of the excitatory and inhibitory potentials and the delay of the inhibition with respect to the excitation determine the model's best modulation frequency (BMF).

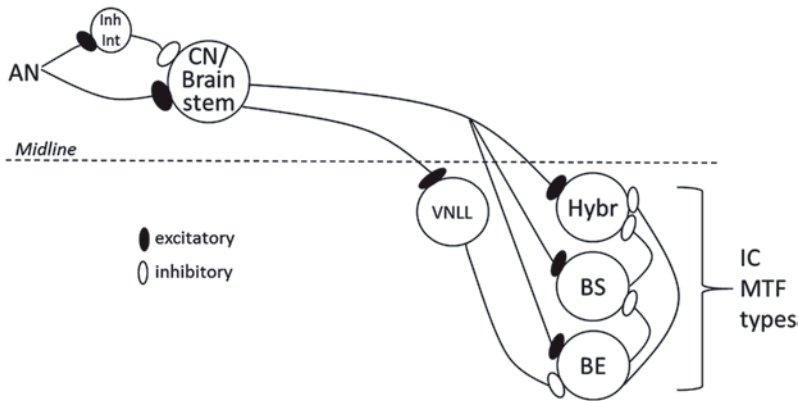
An extension of the SFIE model was used for the band-suppressed model (Fig. 1, BS). This model receives the same ascending excitatory input as the band-enhanced model, but it is inhibited by the band-enhanced model (Carney et al. 2015; Kim et al. 2015b). The BMF of the band-enhanced model determines the worst modulation frequency (WMF) for the band-suppressed model.

Finally, the hybrid MTF model receives the same ascending excitatory input as the other model IC neurons, but is inhibited by both band-enhanced and band-suppressed cells (Fig. 1, Hybr). The relation between the BMF and WMF of the inhibitory inputs determines the shape of the hybrid model cell's MTF.

## 3 Results

The responses of AN models to the vowel/æ/(in "had") are illustrated in Fig. 2, with the vowel waveform (Fig. 2a) and spectrum (Fig. 2b). Discharge rate vs. BF is shown for healthy (Fig. 2c) and impaired (Fig. 2e) model populations. At conversational speech levels, these profiles encode the formant frequencies. However, in background noise or as sound level increases the profiles tend to saturate, though intelligibility does not decrease until high SPLs are reached (Studebaker et al. 1999).

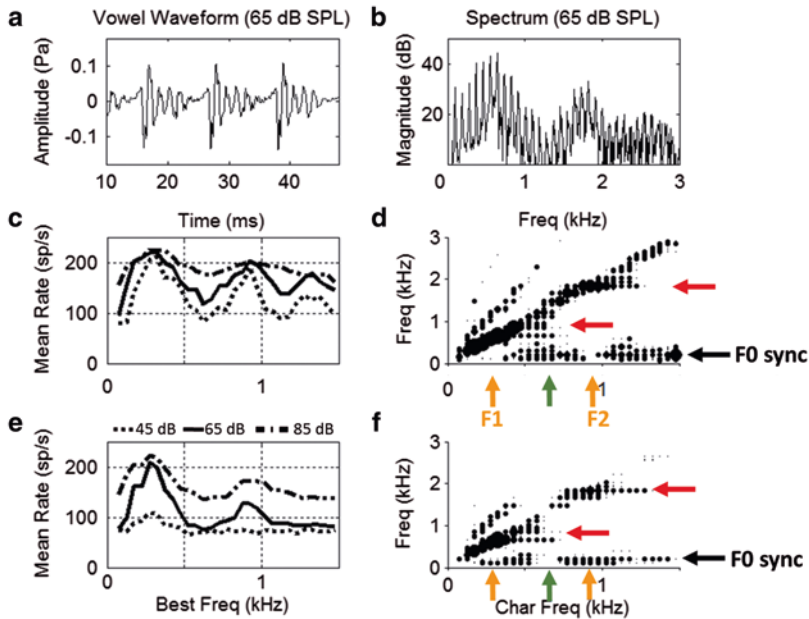
Figure 2d, f illustrates features of the temporal responses of AN fibres to vowels shown using the dominant components, which are the largest spectral components



Synapse	Excitatory	Inhibitory
	Tau (ms), Ampl	Tau (ms), Ampl, Delay (ms)
CN inputs	0.5, 1	2.0, 0.6, 1.0
BE inputs	BE cell: 1.0, 1 BE to BS cell: 2.0, 1	BE cell: 1.0, 1.8, 2.0 BE to BS cell: 6.0, 1.1, 2.0
BS inputs	2.0, 0.8	6.0, 7.0, 2.0
Hybrid inputs	2.0, 0.8	6.0, 8.0 (from BE) & 1.3 (from BS), 2.0

**Fig. 1** Schematic diagram of the extended SFIE model for IC neurons with three types of MTFs: band-enhanced (*BE*), band-suppressed (*BS*) and hybrid. For simplicity, the simulations presented here were driven only by contralateral inputs. The AN excites and inhibits (via an interneuron) a model cochlear nucleus (*CN*) or other brainstem-level neuron, which excites all three IC cell types, and inhibits (via the ventral nucleus of the lateral lemniscus, *VNULL*) the *BE* model. The *BE* model inhibits the *BS* model, and both *BE* and *BS* models inhibit the hybrid model. In each model, inputs were convolved with the post-synaptic potential waveforms, summed, and half-wave rectified

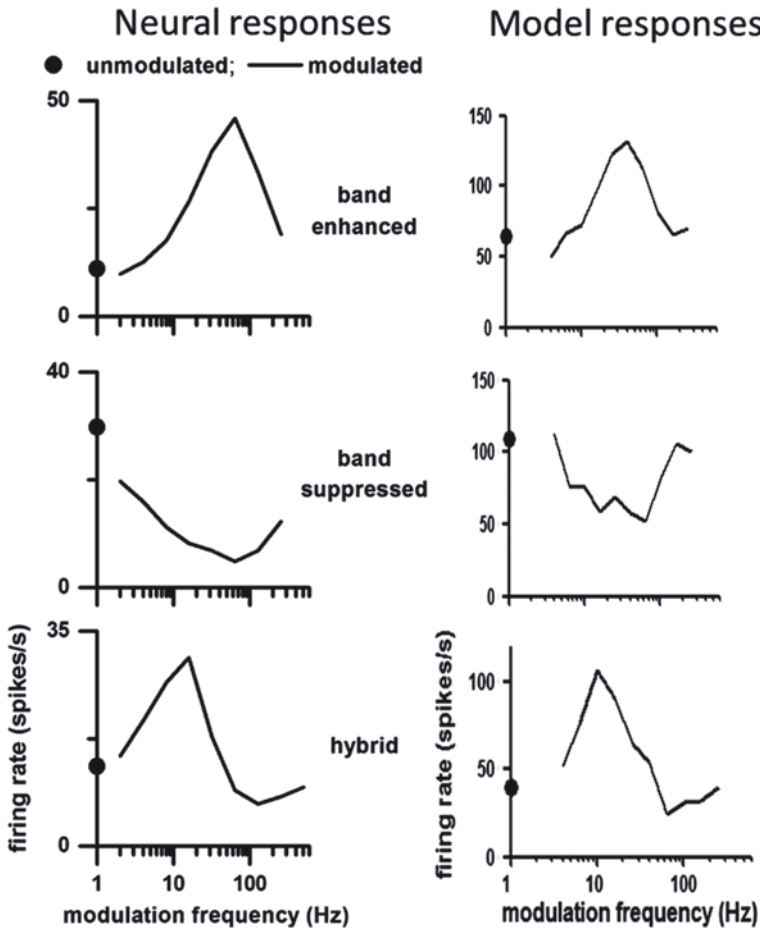
of the post-stimulus time histograms (Delgutte and Kiang 1984; Miller et al. 1997). Dominant components for the healthy AN model (Fig. 2d) show synchrony capture, or dominance of the temporal responses by a single harmonic near the formant peak. The synchrony capture extends across a span of fibres with BFs near each formant frequency (red arrows). The dominance of these responses by a harmonic near the formants (orange arrows) results in absence of synchrony to F0 (black arrow) for BFs near formants. In contrast, fibres tuned between the formants are not dominated by a single harmonic. Because these fibres respond to multiple harmonics near their BFs (green arrow), their responses synchronize to the “beats” at F0 (black arrow).



**Fig. 2** Time waveform (a) and spectrum (b) for the vowel /æ/ (in “had”) at 65 dB SPL (from Hillenbrand et al. 1995; F0=95 Hz). (c) Healthy and (e) impaired AN model rate profiles for 3 sound levels. (d, f) Dominant components of synchronized rate for healthy and impaired responses. Symbol sizes are proportional to synchronized rate (components smaller than 15 sp/sec are not shown)

The dominant components for model AN fibres with SNHL simulated by outer and inner hair cell dysfunction are shown in Fig. 2f. Physiological studies have shown that synchrony capture is weaker in these fibres (Miller et al. 1997). The model response illustrates reduced synchrony capture, especially to F2 (Fig. 2f, red arrow), even for the mild hearing loss in this simulation. The reduced synchrony capture is more evident in the pattern of synchrony to F0, which extends across all regions where there is sufficient energy to drive the neurons (Fig. 2f, black arrow). Thus, the profile of F0-synchronized rates differs qualitatively between the impaired and the healthy AN populations. For BFs tuned between F1 and F2 (green arrow), the response to F0 (black arrow) is reduced in the impaired nerve, whereas in the BF regions of F1 and of F2 (orange arrows), the response to F0 is increased in the impaired nerve. The net effect of SNHL is a reduction in the contrast of the F0-synchronized rates across BF.

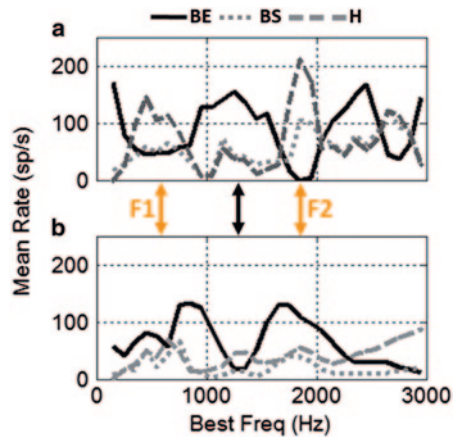
Figure 3 shows examples of IC neural (left) and model (right) band-enhanced, band-suppressed, and hybrid MTFs. Kim et al. (2015a) used multiple types of AM envelopes (e.g., raised-sine envelopes with exponents of 1, 8 and 32) and a 25% criterion for a change in rate with respect to the unmodulated response, and found that approximately 25% of IC neurons had band-enhanced, 50% band-suppressed, and 25% hybrid MTFs.



**Fig. 3** *Left*: Three types of MTFs in the IC of awake rabbit (from Kim et al. 2015a). *Right*: IC model MTFs with BF=2 kHz. All neural and model responses are to 1-octave bandwidth sinusoidal AM noise (100% modulation depth) centred at BF. Filled dot is response rate to unmodulated noise

Figure 4 shows model responses to the vowel /æ/ for the three MTF types for healthy and impaired model AN inputs. The healthy band-enhanced responses (Fig. 4a, blue) reflect the patterns of F0-synchronized rate shown in Fig. 2; responses are decreased near formant frequencies (orange arrows), because synchrony capture in the healthy AN reduces the low-frequency fluctuations that ultimately drive this IC type. Band-enhanced neurons tuned to frequencies between formants (Fig. 4a, green arrow) have strong responses due to the strong F0-synchronized rate in the healthy AN for these frequency channels (Fig. 2d, green arrow). The strong contrast in rate across frequency channels for the band-enhanced responses is robust across SPLs and in background noise (Carney et al. 2015).

**Fig. 4** Responses of model IC MTF types to the vowel/æ/at 65 dB SPL. IC models with (a) healthy AN model inputs, (b) impaired AN inputs. Model parameters provided in Fig. 1



The rate profile for healthy band-suppressed IC model neurons has peaks at the formant frequencies (Fig. 4a, magenta). These cells respond more strongly at frequencies where the low-frequency fluctuations are reduced (Fig. 2d, orange arrows). The peaks in the rate profile at the formant frequencies are robust in the model hybrid population (Fig. 4a, green). This model result motivates further physiological and computational investigation of these cells. The hybrid neuron and model in Fig. 3 are most strongly driven by a contrast in neural fluctuations (i.e., strong neural fluctuations in the 8–30 Hz range and reduced fluctuations in the 30–200 Hz range); other MTF shapes are also observed for hybrid neurons (Krishna and Semple 2000).

Model IC responses with impaired AN inputs (Fig. 4b) differ dramatically from the healthy responses. The band-enhanced rate profile (Fig. 4b, blue) has peaks approximately where the healthy model has valleys, and vice versa. The impaired model profile is explained by reduced synchrony capture: the impaired AN model synchronizes to F0 whenever the fibres are adequately driven by stimulus energy (see Fig. 2f). Thus the impaired band-enhanced rate profile reflects the energy spectrum of the vowel, unlike the healthy rate profile, which is inversely related to energy. The impaired band-suppressed (Fig. 4b, magenta) and hybrid (Fig. 4b, green) model responses also have peaks near F1 and F2. Unlike the healthy case, the inhibition from the shifted peaks in the impaired band-enhanced model suppresses the peaks in the band-suppressed and hybrid models. These results emphasize the fact that the strong rate profiles in the healthy IC model are not simply explained by the rate vs. BF profile of the AN. The healthy band-suppressed and hybrid profiles are created by a synchrony-to-rate transformation between the AN and IC and enhanced by disinhibition at the formant frequencies. In general, the rate profiles for all three response types are qualitatively different for this mildly impaired model due to differences in the F0-synchronized rate between the healthy and impaired AN models (Fig. 2d, f).

## 4 Discussion

Vowel coding studies have historically focussed on representation of spectral peaks, or formants, in AN rates or synchrony to stimulus components near BF. These representations are vulnerable to additive noise and vary with sound level. The sensitivity of midbrain neurons to low-frequency fluctuations inspired a shift of focus to the contrast in pitch-related fluctuations along the BF axis in vowel coding (Carney et al. 2015). In the healthy AN, there is a robust contrast in the profile of the F0-synchronized discharge rate across the BF axis (Fig. 2d). Midbrain neurons' sensitivities to these F0-synchronized rates transforms the temporally coded AN profile into a discharge rate profile (Fig. 4a).

The representation of F0-synchronized rates is affected by several mechanisms that are influenced by SNHL: synchrony capture, threshold (and thus rate saturation), and bandwidth of tuning. Bandwidth, in turn, affects the modulation spectrum of peripheral responses. Wider peripheral bandwidths associated with SNHL result in a wider modulation bandwidth and a reduction of the amplitude of low-frequency fluctuations. These factors have little influence on the mean rates of AN fibres; however, they have a large effect on F0-synchrony in the AN, which in turn affects the responses of IC neurons.

These IC model responses have implications for the design of signal-processing strategies for listeners with SNHL. Recreating stimulus spectra in the AN rate profiles will not elicit appropriate responses from central neurons, whereas restitution of the F0-synchronized rate profiles may (Rao and Carney 2014). This result suggests an experimental question: If the profile of peripheral F0-synchrony in response to voiced sounds can be restored to normal for these listeners, can intelligibility of speech, and especially speech in background noise, be improved? The results of such a test would depend on the status of central auditory pathways, which may undergo change following cochlear trauma (Suneja et al. 1998; Salvi et al. 2000).

**Acknowledgments** Supported by NIDCD-010813.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.



## References

- Carney LH, Li T, McDonough JM (2015) Speech coding in the brain: representation of formants by midbrain neurons tuned to sound fluctuations. *eNeuro* 2(4) e0004-15.2015 1–1. doi: 10.1523/ENEURO.0004-15.2015
- Delgutte B, Kiang NY (1984) Speech coding in the auditory nerve: i. Vowel-like sounds. *J Acoust Soc Am* 75:866–878
- Deng L, Geisler CD (1987) Responses of auditory-nerve fibers to nasal consonant–vowel syllables. *J Acoust Soc Am* 82:1977–1988
- Deng L, Geisler CD, Greenberg S (1987) Responses of auditory-nerve fibers to multitone complexes. *J Acoust Soc Am* 82:1989–2000
- Fant G (1960) Acoustic theory of speech production. Mouton, Hague
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K (1995) Acoustic characteristics of American English vowels. *J Acoust Soc Am* 97:3099–3111
- Kewley-Port D, Burkle TZ, Lee JH (2007) Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J Acoust Soc Am* 122:2365–2375
- Kim DO, Bishop BB, Kuwada S, Carney LH (2015a). Band-Enhanced and Band-Suppressed Rate Modulation Transfer Functions of Inferior Colliculus Neurons and a Model: effects of Duty Cycle and Rise/Fall Rate, Association for Research in Otolaryngology. 38th Annual Midwinter Meeting, abstract #PS-423
- Kim DO, Zahorik P, Carney LH, Bishop BB, Kuwada S (2015b) Auditory distance coding in rabbit midbrain neurons and human perception: monaural amplitude modulation depth as a cue. *J Neurosci* 35(13):5360–5372
- Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* 84:255–273
- Langner G (1992) A review: periodicity coding in the auditory system. *Hear Res* 60:115–142
- Miller RL, Schilling JR, Franck KR, Young ED (1997) Effects of acoustic trauma on the representation of the vowel/e/in cat auditory-nerve fibers. *J Acoust Soc Am* 101:3602–3616
- Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J Acoust Soc Am* 116:2173–2186
- Rao A, Carney LH (2014) Speech enhancement for listeners with hearing loss based on a model for vowel coding in the auditory midbrain. *IEEE Trans Bio-med Eng* 61:2081–2091
- Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am* 66:470–479
- Salvi RJ, Wang J, Ding D (2000) Auditory plasticity and hyperactivity following cochlear damage. *Hear Res* 147(1):261–274
- Studebaker GA, Sherbecoe RL, McDaniel DM, Gwaltney CA (1999) Monosyllabic word recognition at higher-than-normal speech and noise levels. *J Acoust Soc Am* 105:2431–2444
- Suneja SK, Potashner SJ, Benson CG (1998) Plastic changes in glycine and GABA release and uptake in adult brain stem auditory nuclei after unilateral middle ear ossicle removal and cochlear ablation. *Exp Neurol* 151(2):273–288
- Young ED (2008) Neural representation of spectral and temporal information in speech. *Philos Royal Soc London B: Biol Sci* 363:923–945
- Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc Am* 66:1381–1403
- Zilany MS, Bruce IC (2007) Representation of the vowel/in normal and impaired auditory nerve fibers: model predictions of responses in cats. *J Acoust Soc Am* 122(1):402–417
- Zilany MSA, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. *J Acoust Soc Am* 135:283–286