Effects of sensorineural hearing loss on formant-frequency discrimination: Measurements and models

Laurel H. Carney[1,2], David A. Cameron[1], Kameron B. Kinast[3], C. Evelyn Feld[2], Douglas M. Schwarz[1,2], U-Cheng Leong[4], Joyce M. McDonough[5]

[1]Department of Biomedical Engineering, University of Rochester
[2]Department of Neuroscience, University of Rochester Medical Center
[3]Bridges-to-the-Doctorate Program, Rochester Institute of Technology
[4]Department of Otolaryngology, University of Rochester Medical Center
[5]Department of Linguistics, University of Rochester

ABSTRACT

In the response of the healthy ear to a harmonic sound, auditory-nerve rate functions fluctuate at the fundamental frequency, f0. The amplitude of these fluctuations varies along the tonotopic axis. Neural fluctuation (NF) amplitudes encode spectral peaks, including formant frequencies of vowels, because responses of inner-hair-cells (IHCs) tuned near spectral peaks are captured (or dominated) by a single harmonic, resulting in lower fluctuation amplitudes than responses of IHCs tuned between spectral peaks. This NF code is robust across a wide range of sound levels and in background noise. The NF profile is converted into a rate-place representation in the auditory midbrain, wherein neurons are sensitive to low-frequency fluctuations. The NF code is vulnerable to sensorineural hearing loss (SNHL) because capture depends upon saturation of IHCs, and thus the interaction of cochlear gain with IHC transduction. In this study, formant-frequency discrimination limens (DL_{FF}s) were estimated for listeners with normal hearing or mild to moderate SNHL. The difficulty of the task was modulated by varying formant bandwidth to modulate the contrast in the NF profile. The formants were either aligned with harmonic frequencies or placed between harmonics. Formant frequencies were 600 and 2000 Hz, in the range of first and second formants of several vowels. Results were compared to predictions from model auditory-nerve (AN) and inferior colliculus (IC) neurons, with listeners audiograms used to individualize the AN model. Correlations between DL_{FF}s, audiometric thresholds near the formant frequencies, age, and scores on the Quick speech-in-noise test are reported. SNHL had a strong effect on DL_{FF} for the second formant frequency (F2), but relatively small effect on DL_{FF} for the first formant. The IC model provided good predictions of elevated listener thresholds to F2 as a function of SNHL.

INTROUCTION

The importance of speech as a communication signal motivates studies of its neural encoding. In listeners with normal hearing, speech intelligibility is robust across a wide range of sound levels and background noise, and in the presence of temporal and spectral distortions. In contrast, listeners with even relatively small amounts of hearing loss have difficulty understanding speech in noise. A better understanding of neural speech coding would illuminate this difficulty and improve signal-processing strategies designed to aid these listeners. This study focused on perception of vowel-like sounds in

listeners with normal hearing or with a range of sensorineural hearing loss (SNHL) to test the hypothesis that discrimination of spectral peaks, or formants, is consistent with a code based on the profile of neural fluctuations (NFs) in auditory-nerve (AN) responses.

In response to voiced sounds, including vowels, most AN responses are dominated by temporal phase-locking to either the fundamental frequency, f0, or to harmonic frequencies near the fiber's characteristic frequency (CF, the frequency at which a fiber is most sensitive) (Young & Sachs, 1979; Delgutte & Kiang, 1984). The dominance of f0 wanes for fibers tuned near spectral peaks, for which responses are "captured," or dominated, by phase-locking to the harmonic closest to CF (Deng and Geisler, 1987; Delgutte & Kiang, 1984; Young & Sachs, 1979; Zilany and Bruce, 2007). Early studies focused on the possibility that this phase-locking to the temporal fine structure (TFS) of the harmonics could provide a code for spectral peaks, yet identification of neural mechanisms that decode TFS phase-locking has been elusive. The NF model focuses instead on the amplitude of the f0-related fluctuations in the AN responses (Carney et al., 2015; Carney & McDonough, 2019; Carney, 2018). These low-frequency fluctuations are important for driving response at the level of the auditory midbrain, or inferior colliculus (IC), where nearly all neurons are sensitive to low-frequency fluctuations of their inputs (Krishna and Semple, 2000; Nelson and Carney, 2007; review: Joris et al., 2004). The decoding of the NF profile is straightforward: average discharge rates across the population of IC neurons represent the NF profile across the population of AN fibers. These response characteristics along the ascending auditory pathway are illustrated in Fig. 1.

AN fibers tuned away from spectral peaks have large amplitude NFs at f0, due to beating between harmonics in the inner ear that is transduced by inner hair cells (IHCs). However, near spectral peaks, IHC responses saturate, and as a result the f0-related fluctuations are "flattened," reducing the amplitude of the NFs in responses of AN fibers tuned near spectral peaks. The profile of NF amplitudes along the tonotopic axis thus encodes the locations of spectral peaks (Carney et al., 2015). Information about the frequencies of the spectral peaks, or formant frequencies, allow identification of vowels. More generally, NFs encode both voiced and unvoiced speech sounds (*e.g.*, fricative consonants, Hamza et al., in review).

NFs depend on IHC sensitivity and saturation, and thus are directly influenced by cochlear gain. As a result, the reduced cochlear gain associated with SNHL affects any neural code based upon NFs. Here we tested the hypothesis that NFs encode the spectral peaks of vowel-like sounds. The bandwidths of formants were manipulated to increase or reduce the contrast in the NF profiles. We also estimated formant-frequency discrimination thresholds in listeners with a wide range of hearing sensitivity. We hypothesized that thresholds for all listeners would increase (worsen) as formant bandwidth was increased, and that thresholds would increase proportionally to hearing loss due to SNHL-dependent reduction in the contrast of NF profiles. Lastly, for all listeners and formant bandwidths, we varied the frequency of the formant peak with respect to the harmonic frequencies, a manipulation previously shown to affect formant-frequency discrimination (Lyzenga & Horst, 1995, 1997; Tan & Carney, 2006; Henry et al., 2017).

Our findings suggest that discrimination of F2 is more vulnerable to hearing loss than is F1. This finding has implications for the effect of hearing loss on speech understanding. The role of formant-frequency discrimination for F1 and F2 in determining the systematic patterns of vowel contrasts cross-linguistically will be discussed.

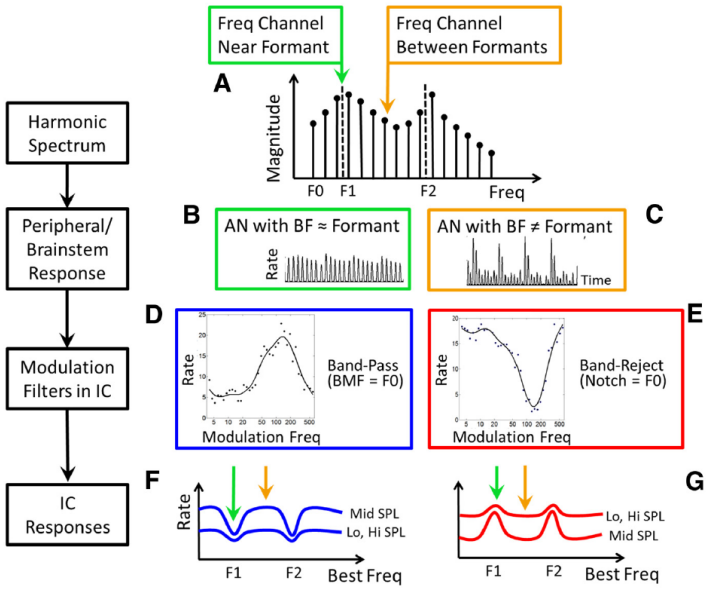


Figure 1 (from Carney et al., 2015 paper) (see Carney et al., 2016)

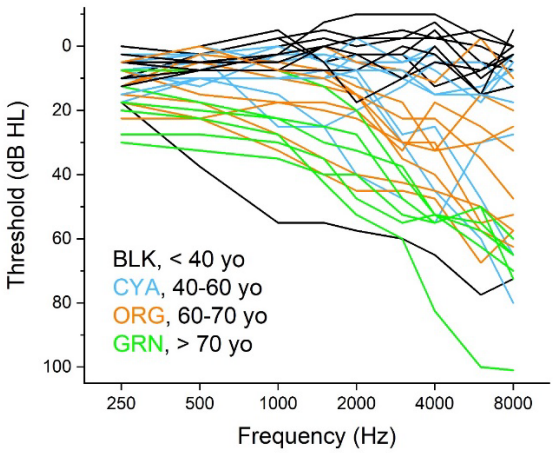
METHODS

Listeners. Thirty-four participants (ages 18-79 years, 26 female, 8 male) were recruited and tested using procedures approved by the University of Rochester Institutional Review Board. All participants were initially screened with a standard audiogram (0.25 – 8 kHz; Fig. 2) and were excluded if hearing loss (HL) was asymmetric (> 15 dB difference in HL between the two ears) or if HL was greater than 80 dB. Audiometric thresholds were used as input parameters for the computational models (see below), and thresholds at 500 Hz and 2 kHz was used in analyses of formant-frequency discrimination thresholds. All native English-speaking subjects (*n* = 31) were screened with the Quick Speech-in-Noise test (QSIN; Killion et al., 2004). Individual test sessions were one or two hours in duration.

Stimuli. Vowel-like stimuli were synthesized based on Klatt (1980), using the glottal pulse of Rosenberg (1971). Stimuli had a fundamental frequency (F0) of 100 Hz, duration of 300 ms, 25 ms $\cos^2$ on/off ramps, and level of 75 dB SPL. DL_{FF}s were estimated for synthetic vowels with two formant confgurations; the formants were either aligned with harmonics (ON; F1=600 Hz, F2 = 2000 Hz) or midway between harmonics (BTW; F1 = 650 Hz, F2 = 2050 Hz) (Fig. 3). DL_{FF}s were also estimated for three formant bandwidths (50, 100, and 200 Hz; Fig. 3). The 12 stimulus conditions were tested in blocks that were randomly ordered for each subject.

Procedure. A two-interval forced choice, two-down one-up tracking procedure was used, as follows. Each trial contained two pairs of sounds; the standard pair had identical sounds, and the target pair had sounds that differed in either F1 or F2. For each track the formant-frequency difference (ΔF) started at 40%. The formant-frequency difference was reduced by a factor of two following every two correct

responses, or increased after incorrect responses, until two reversals occurred. ΔF was then reduced or increased by a factor of 1.26 until four reversals occurred, then by reduced or increased by a factor of 1.12 until a total of ten track reversals were attained. Individual tracks were retained for analysis if a criterion of (standard deviation of ΔF / track threshold) < 0.3 was achieved. The DL_{FF} for each 10-reversal track was estimated as the average ΔF of the final six reversals. Overall threshold for a particular stimulus condition was based on the average of at least three individual track thresholds. If the ratio of the standard deviation of the threshold estimates to the mean threshold was > 0.3, additional tracks were collected until the criterion (< 0.3) was achieved. Statistical analyses were based on ANOVA analyses of linear mixed-effects models, focusing on the bandwidth factor and its interaction with the factors of HL, QSIN, and Age.



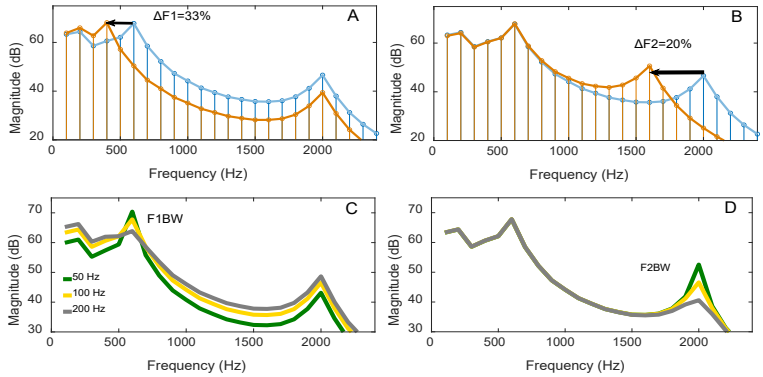


Fig 2. Subject audiograms. Audiograms (mean of both ears) for each subject. Age ranges are indicated by line color (see Fig. 7 for individual subject ages).

Fig. 3 Representation of stimulus spectra. **(A, B)** Discrimination thresholds were measured using a standard pair of identical sounds with formant peaks centered on (ON) harmonics of F0: F1=600 Hz and F2=2000 Hz (Blue). One sound in the target interval differed in either F1 (A) or F2 (B) (Orange).

Thresholds were also measured for formants between (BTW) two harmonics of F0 (F1=650 Hz and F2=2050 Hz; stimuli not shown). (**C, D)** Formant bandwidths used for F1 (C) and F2 (D). Bandwidths = 50 (green), 100 (yellow), and 200 (gray) Hz.

Computational Modeling Methods

Model thresholds were estimated as follows: Auditory-nerve (AN) responses were simulated for 11 high-spontaneous-rate fibers with log-spaced CFs spanning 1/3-octave, centered at the formant frequency being tested. Individual audiograms were included in the AN model (Zilany et al., 2014).

Model DL_{FF}s were based on Band-Suppressed (BS) IC model neurons (Carney & McDonough, 2019). The BS population rate profile peaks at frequencies near formants, where fluctuations are reduced by capture (Fig. 1). Midbrain neurons are either excited or suppressed by temporal fluctuations on their inputs. Fluctuations are minimal near formant peaks, due to capture of the IHC responses by a single harmonic near CF (Deng and Geisler, 1987). BS models had a trough frequency at 64 Hz, the median for rabbit IC neurons (Kim et al., 2020).

The stimulus interval with model population responses having the largest mean squared difference across the pair of stimuli was chosen as the target interval. Each DL_{FF} estimate was based on a logistic function fitted to %-Correct vs. ΔF.

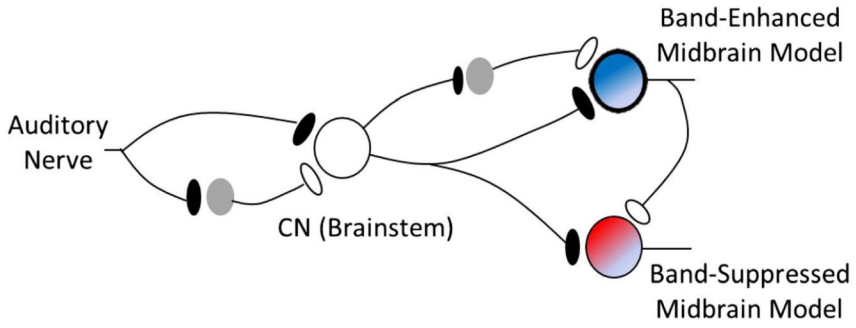


Fig. 4 – Schematic illustration of model IC neuron (from Carney et al., 2015)

RESULTS

Thresholds vs. HL at formant frequencies

Subject thresholds for the F2 and F1 conditions, as a function of HL and stimulus bandwidth, are presented in Fig. 5; results from ANOVA analysis of linear mixed-effects models for the same data are presented in Table 1. For the F2 condition there was a strong, positive, linear correlation between threshold and HL. These correlations were similar for all three tested bandwidths, and also for both ON and BTW stimuli (see Fig 5 legend for linear regression R^2 values). A significant bandwidth-dependent increase in threshold was also observed for both ON and BTW, being particularly evident for subjects with HL < 15 dB (*cf*. Fig. 5 top row panels), and confirmed with ANOVA analysis (Table 1). For the F1

condition there was little evidence for positive correlation between threshold and HL for any of the tested bandwidths (Fig. 5, bottom row; Table 1). Similar to the F2 condition, a significant bandwidth-dependent increase in threshold was observed in the F1 condition for both ON and BTW, particularly evident for subjects with HL <15 dB, and was confirmed by ANOVA analysis (Table 1). These results are consistent with a statistically significant, positive, bandwidth- and mid-to-high frequency HL-dependent difficulty in discriminating differences in F2, a difficulty that is comparatively absent for F1.

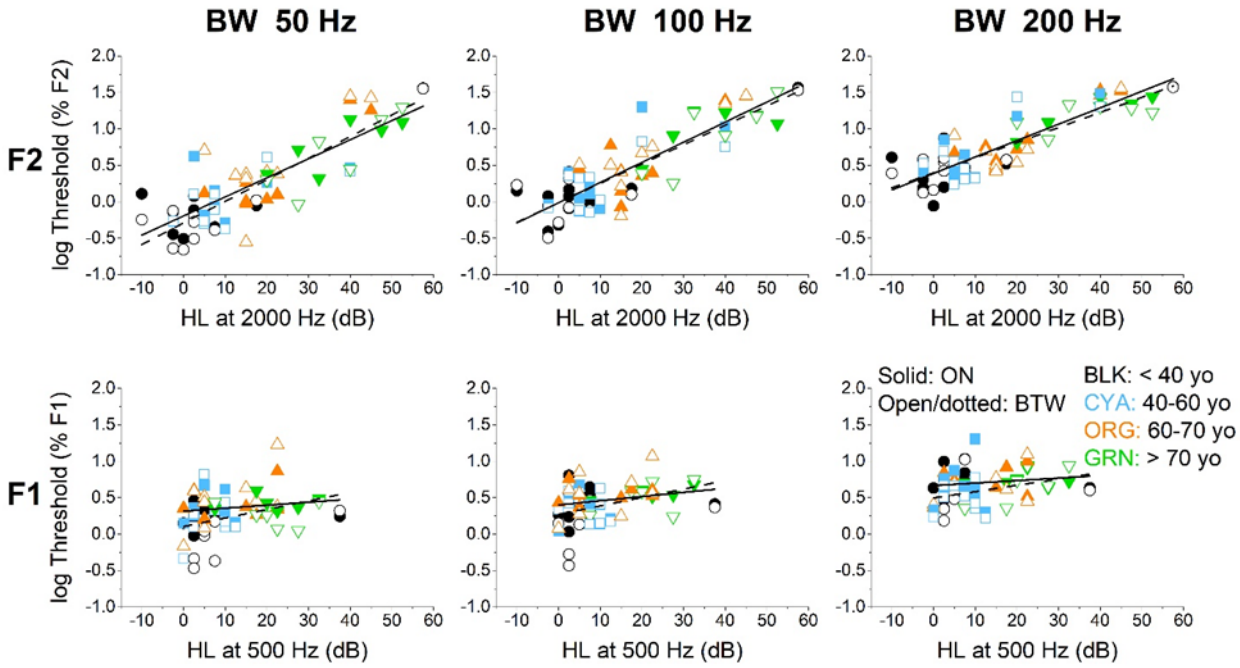


Figure 5 F2 and F1 threshold as a function of HL and stimulus bandwidth. For each plot the aggregate data for ON (solid symbols) and BTW (open symbols) are fitted with linear regressions (solid and dotted lines, respectively); data points are color- and symbol-type-coded by age range as indicated (see also Fig, 7). For the F2 condition (top row), thresholds for both ON and BTW stimuli increased significantly with increasing HL at 2000 Hz (linear regression R2 values: ON, 0.72 (50 Hz), 0.74 (100 Hz), 0.75 (200 Hz); BTW, 0.70 (50 Hz), 0.72 (100 Hz), 0.71 (200 Hz)). For the F1 condition (bottom row) there was little evidence for correlation between threshold and HL at 500 Hz (linear regression R2 values: ON, 0.02 (50 Hz), 0.06 (100 Hz), 0.01 (200 Hz); BTW, 0.08 (50 Hz), 0.11 (100 Hz), 0.11 (200 Hz)). In both the F2 and F1 conditions, for HL < 15 dB elevated thresholds for both ON and BTW was evident for BW 200 Hz compared to the narrower bandwidths. See Table 1 for ANOVA analyses of the data plotted here.

Ratiometric comparison of ON and BTW thresholds as a function of HL and stimulus bandwidth are presented in Fig. 6. For subjects with HL <15 dB at either 2000 or 500 Hz, thresholds for ON were typically higher than BTW for both the F2 and F1 conditions, particularly for subjects less than age 40 (black symbols, Fig. 6). This trend was reduced for all subjects, and for both the F2 and F1 conditions, as a function of increasing stimulus bandwidth. These results indicate for most subjects a general, but not

absolute, greater sensitivity to differences in BTW stimuli compared to differences in ON sounds, for both the F2 and F1 components of artificial vowel sounds.

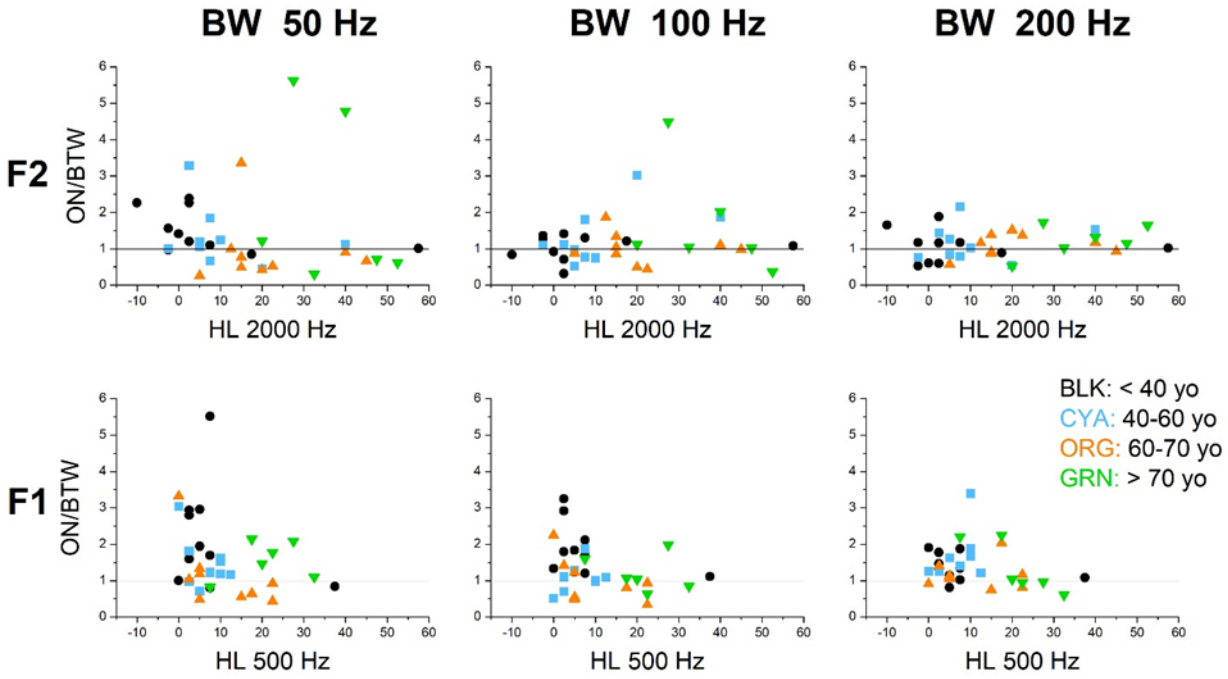


Figure 6. Ratio of ON and BTW thresholds as a function of HL and bandwidth. For F2 (top row), normal-hearing subjects (<15 dB HL) tend to have greater thresholds for ON than BTW at 50 Hz bandwidth, a trend that is absent at greater bandwidths; subjects with mild-to-moderate HL tend to have greater threshold for ON compared to BTW as a function of increasing bandwidth. For F1 (bottom row), normal-hearing subjects have a consistently higher threshold for ON than BTW, independent of HL and bandwidth; the thresholds for ON and BTW are similar with increases in HL and bandwidth.

QSIN scores as a function of subject age and HL are presented in Fig. 7. The strongest linear correlation (*i.e.*, regression R^2 value) was observed for HL at 2000 Hz (Fig. 7, right panel). This result was supported by ANOVA analysis (Table 1). Two outlier QSIN scores > 10 for two subjects suppressed these correlations, particularly with respect to HL; when those two data points are excluded the R^2 values are 0.17 (age), 0.38 (HL 500 Hz), and 0.50 (HL 2000 Hz). These results indicate a general, positive correlation between QSIN score and HL, and further indicate that formant frequency discrimination tasks, such as those used in this study, can be used as "non-cognitive" surrogates to the QSIN test to assess HL and vowel-sound discrimination in both native and non-native English speakers.

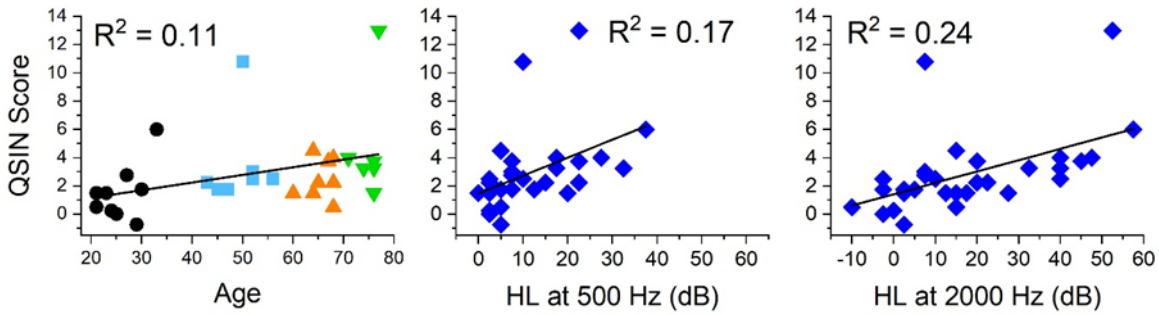


Figure 7. QSIN scores as a function of age and HL. Linear regression R2 values are indicated; color and symbol-type (left panel) match that of Figs. 5 and 6. There were weak correlations between QSIN score with age and HL, particularly for HL at 2000 Hz (right panel). See Table 1 for ANOVA analyses of these data.

Statistical Analysis

Statistical evaluation (P-values of ANOVA analysis of linear mixed-effects models) of the human-derived thresholds is summarized in Table 1. For both F1 and F2 in both the ON and BTW conditions, bandwidth had a significant effect upon DL_{FF}, with no significant interaction between bandwidth and age. There was a significant effect of age (green box, Table 1) and hearing loss upon the F1 BTW condition, for which some young subjects with low audiometric thresholds had especially low thresholds (presented graphically below). For F2, but not F1, QSIN score, age, and hearing loss at 2 kHz all had significant effects on DL_{FF}s for both the ON and BTW conditions. Interactions between bandwidth and hearing loss, and between bandwidth and QSIN score, were significant only for the F2 BTW condition. The effects of hearing loss near the formant frequency, QSIN score, and age were always significant for the F2 thresholds, the only significant interaction for the F1 BTW condition being that between age and bandwidth. These statistical analyses indicated that all interactions between a factor (age, QSIN score, and hearing loss) and bandwidth were significant for the F2 BTW condition.

Factors and Interactions	F1 ON	F1 BTW	F2 ON	F2 BTW
Bandwidth	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
HL @ 500 (F1) or 2000 (F2) (N = 34)	0.20	0.02	$p < 0.001$	$p < 0.001$
Bandwidth	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
HL:Bandwidth	0.68	0.75	0.08	0.004
QSIN scores (N = 31)	0.48	0.18	0.013	0.009
Bandwidth	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
QSIN:Bandwidth	0.16	0.34	0.64	0.015
Age (N=34)	0.14	0.007	0.001	$p < 0.001$
Bandwidth	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Age:Bandwidth	0.47	0.02	0.20	0.02

Table 1. P-values of ANOVA analysis of Linear Mixed-Effects Models, focusing on the bandwidth factor and its interaction with the factors of hearing loss (HL), QSIN score, and age.

Model IC BS population responses to four stimulus pairs are illustrated in Fig. 8. Square differences between these IC rate profiles were used as the decision variable to estimate model thresholds, using the same 2-pair (i.e., 4-interval) stimuli for each trial as was used in listening tests. The differences in the IC BS rate profiles between the ON and BTW conditions is due to differences in the nature of the stimulus fluctuations when the formant is aligned with a harmonic as opposed to when it falls between two harmonics (Lyzenga and Horst, 1997; Tan and Carney, 2006; Henry et al., 2017).

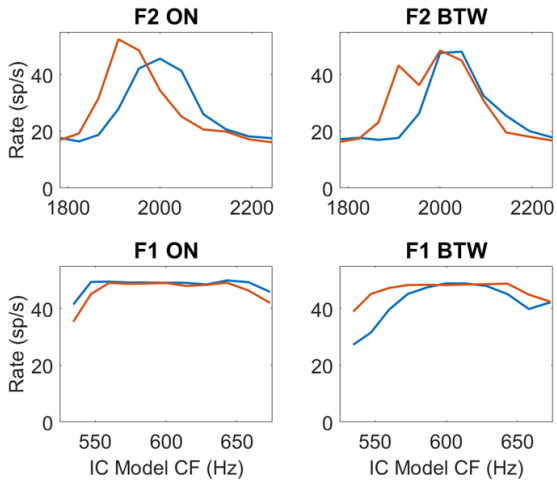


Fig. 8 IC model Profiles. IC model rate profiles for stimuli with ΔF = 3%, BW = 100 Hz. Blue = Standard stimulus; Red = Formant shifted by 0.01(ΔF) × F0. Threshold estimates were based on the sum of squared errors between these profiles.

The AN model results as a function of subject thresholds are presented in Fig. 9. In the F2 condition, for bandwidths of 100 and 200 Hz the model-derived thresholds displayed some positive correlation with the subject thresholds, and roughly captured the human threshold values (R^2 range of 0.28-0.36 for ON and BTW). Such correspondence was largely absent for the 50 Hz bandwidth condition, with the model-derived threshold range being less than that of the human threshold range. For F1 and for all three bandwidths, the AN model consistently returned elevated thresholds (points above the unity line on Fig. 9) that were mostly negatively correlated with the human threshold values. The AN model results did qualitatively resemble the 'clumped' nature of the human results in the 200 Hz bandwidth condition. These model results suggest that the AN model captures some aspects of human F2 and F1 thresholds, but that absolute model threshold values are consistently higher for the F1 condition.

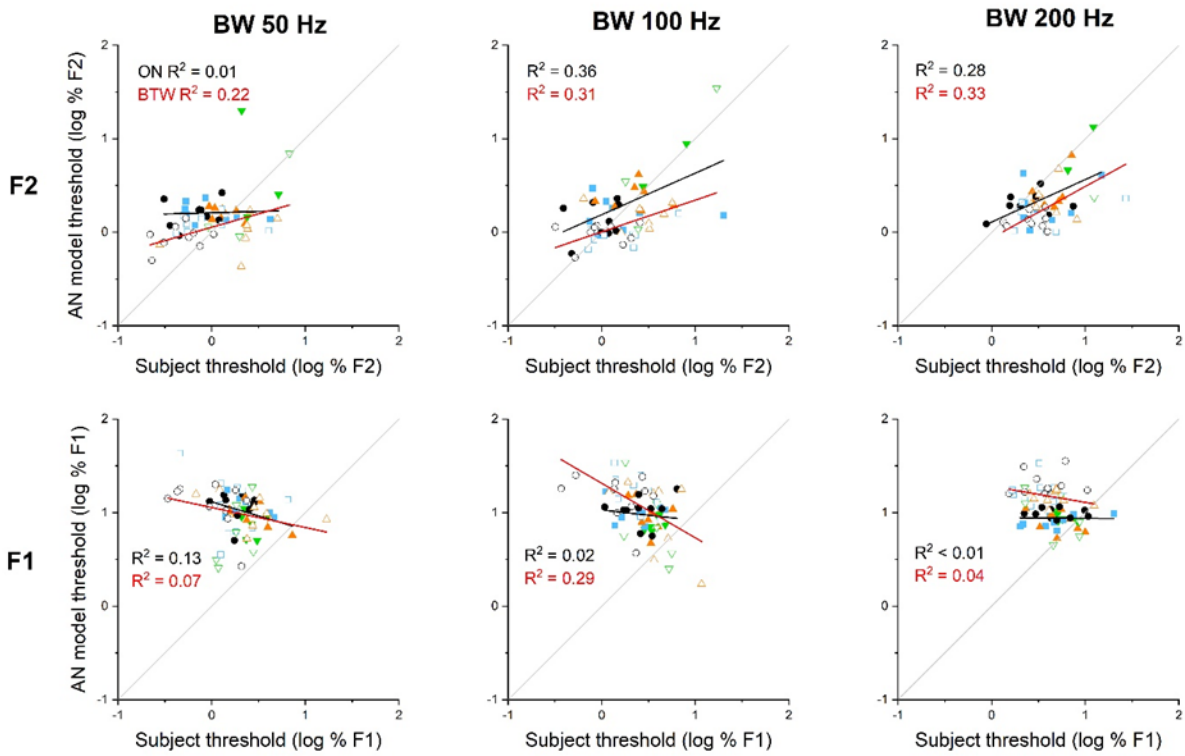


Figure 9. AN model thresholds as a function of subject thresholds. Symbol types and colors as in Fig. 5, with the gray line indicating unity. Bold lines correspond to linear regressions of the displayed ON (black) and BTW (red) data sets, with the R2 values indicated in each panel. The AN model captured some of the F2 data trends for 100 and 200 Hz bandwidths, but less so for the 50 Hz bandwidth (cf. the respective panels in Fig. 5). For F1 the AN model consistently had elevated thresholds (points above the unity line) that were mostly negatively correlated with the human listener values.

The IC model results as a function of subject thresholds are presented in Fig. 10. Compared to the AN model, the IC model returned better estimates of human thresholds for the F2 condition, being particularly well-correlated for the 100 and 200 Hz bandwidth conditions (with correspondingly greater R^2 regression values; top row, Fig. 10). These F2 estimates were also well-matched to the absolute threshold values across the entire range of subject thresholds, although the IC model results for 100 and 200 Hz tended to slightly underestimate the human thresholds (*i.e.*, regression lines just below the unity line). The IC model also performed better than the AN model in estimating the subjects' F1 thresholds (*cf*. linear regressions in the bottom rows of Figs. 9 and 10). For the F1 ON condition the IC model estimated well both the 'clumped' nature, and absolute threshold values, of the human thresholds for the ON condition. For the F1 BTW condition, however, with the exception of subjects with the lowest thresholds, the IC model consistently returned threshold values substantially lower than the corresponding human values (*i.e.*, points and red regression lines below the gray unity line in Fig. 10, bottom row), a trend that decreased with increasing bandwidth. These results indicate that detection cues available to the IC model might also be used by human subjects, especially for the F2 condition at

100 and 200 Hz (both ON and BTW). For the F1 condition, the model apparently takes advantage of enhanced fluctuation cues present in the BTW condition, with only the most sensitive human subjects apparently using those cues.

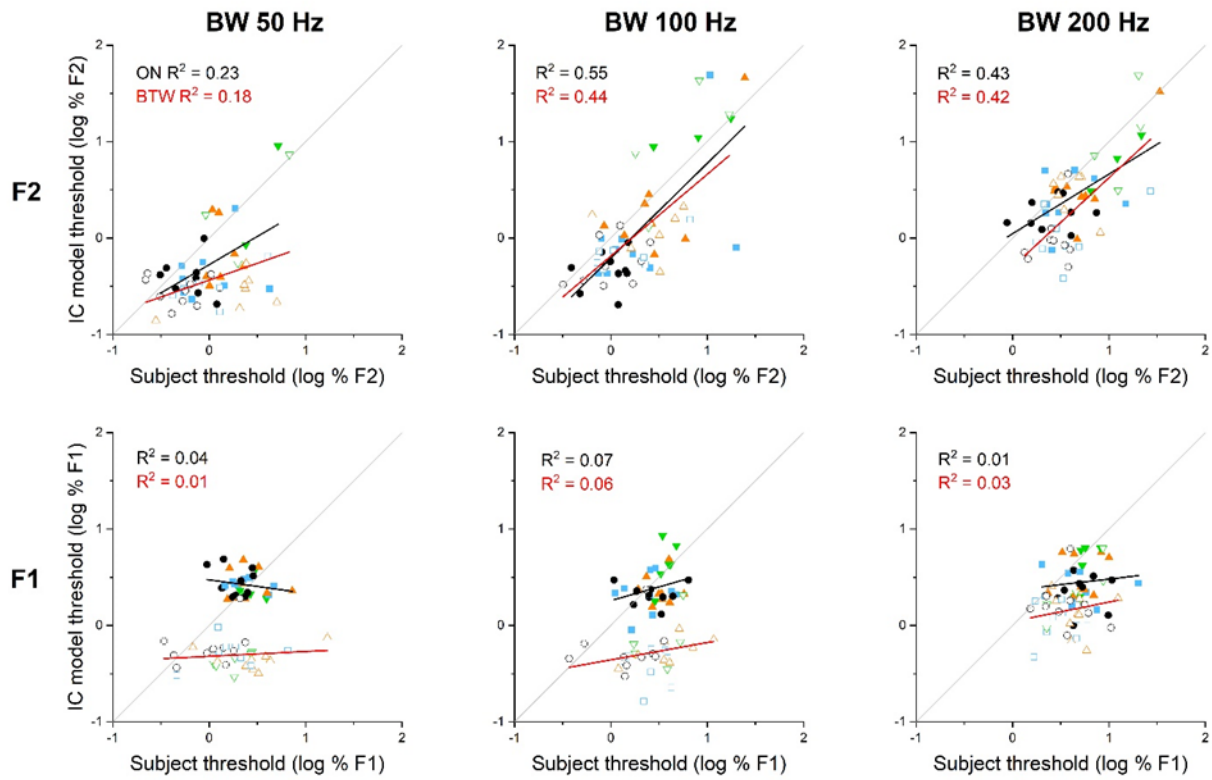


Figure 10. IC model thresholds as a function of subject thresholds. Symbol types, colors, and lines as in Fig. 9. Compared to the AN model, the IC model better represented ON and BTW human subject data for the F2 condition at all three bandwidths (*cf*. corresponding panels in Fig. 9). For F1, the IC model generally matched the values and clumped nature of the ON subject data at all three bandwidths, but consistently had lower thresholds than the human data for the BTW condition (open symbols, bottom row).

DISCUSSION

Formant-frequent discrimination increased systematically with formant bandwidth, for both F1 and F2, for ON and BTW conditions. This trend was present in subject data and in IC model thresholds.

For **F2, formant-frequency discrimination thresholds were correlated with hearing loss at 2000 Hz,** for both subjects and the IC model.

For **F1, thresholds did not vary significantly with hearing loss at 500 Hz,** for both subjects and the IC model.

The **largest discrepancy between model and subject thresholds was for F1 BTW condition**. The model thresholds were generally lower than subject thresholds – except for the most sensitive subjects. **These subjects, and the IC model, may have taken advantage of strong fluctuation cues present in the BTW condition** (Henry et al., 2017; Lyzenga & Horst, 1995).

QSIN scores were correlated to thresholds for F2, most strongly for the F2 BTW condition. Thus, this condition may provide **an alternative to intelligibility testing for quantifying sensitivity of individuals to changes in speech sounds**.

Future studies will explore strategies for enhancing intelligibility by manipulating neural-fluctuation profiles. The potential implication of an elevated FFDL for F2 on intelligibility is illustrated by the loss of resolution on the F2 axis of the English-language vowel space (Fig. 11).

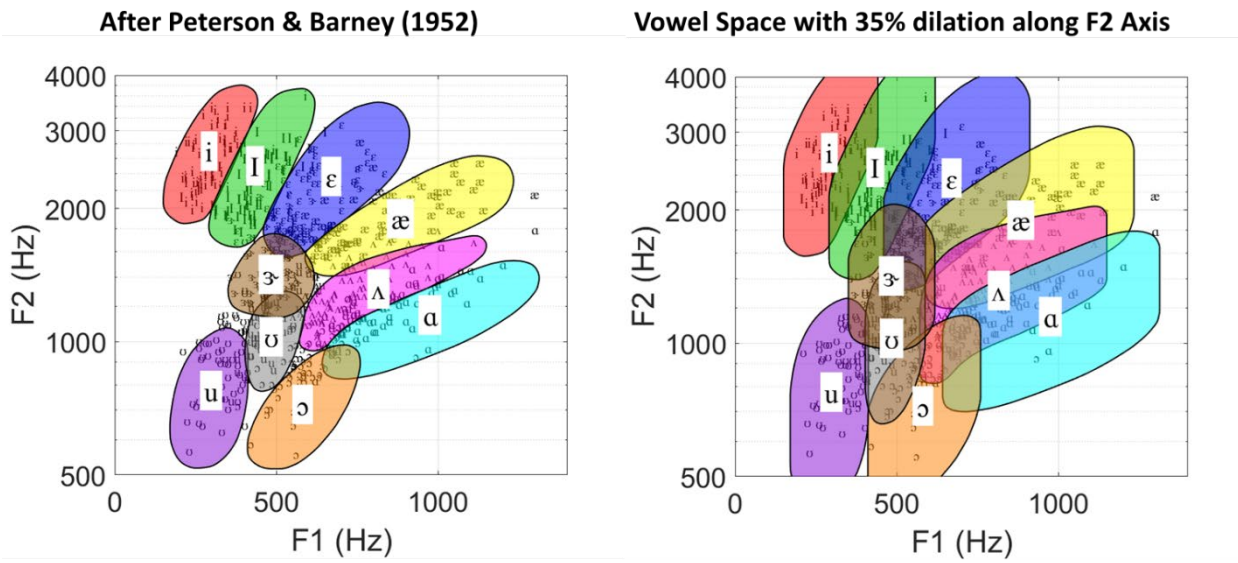


Fig. 11. – Vowel space (A) with 'dilation' corresponding to 35% increase in F2 range (+/- 17% w.r.t. P&B) based on changes in DL_{FF}s for listeners with SNHL. Schematic visualization of the implications of elevated FFDLs for vowel discrimination. NOTE: Threshold for listeners with most hearing loss in F2, BW=100 Hz condition was 35% [log10(35%) = 1.54 on log axis] for both ON and BTW conditions. This listener required a 35% difference in the frequency of F2 to reliably detect a change, in comparison to subjects with 0 dB HL who required ~1% difference. Diagram shows schematically the increase in overlap of the vowels in the vowel space that might correspond to higher DL_{FF}s for F2. This overlap would predict HL-dependent vowel confusions (is there any confusion data in the literature that this result would help explain?).

For our listeners, F1 DL_{FF}s were not significantly elevated by HL at 500 Hz, thus one would expect that their "resolution" along the F1 axis of a 2D vowel space would not be affected. However, F2 DL_{FF}s were significantly elevated by HL at 2 kHz, by as much as 35% for listeners in this study with the largest HL. Figure 11 shows Peterson & Barney (1952) vowel space for English to illustrate the effect of a 35% decrease in F2 resolution, i.e., a 35% increase in the size of vowel clusters. Considerable overlap between vowel clusters is consistent with elevated QSIN scores for listeners with high F2 DL_{FF}s.

These results shed new light on the structure of vowel contrasts and systems. In a well-documented, cross-linguistic phenomenon, observed vowel contrasts pattern in consistent ways, dispersing themselves within a space defined by the acoustic parameters of the vowels' first two formants, F1 and F2, called Dispersion Theory (Liljencrants & Lindblom, 1972; Lindblom 1986; Schwartz et. al. 1997a,b; Becker-Kristal, 2010). A given vowel space is defined by the highest front vowel [i] and the lowest vowel [a] (Becker-Kristal, 2010), with larger systems occupying a larger space. However, the larger vowel systems differ from smaller systems in the number of vowel contrasts, mainly along the F1 (height) parameter. Attempts at modeling the vowel space overpredict the number of vowel contrasts that appear along the F2 (backness) parameter (Liljencrants & Lindblom, 1972; Schwartz et al., 1997a). The acoustics cannot explain the vowel-dispersion patterns, and nearly every study in this area has called on the need to understand the constraints of the auditory system on vowel patterns. Our finding that formant-frequency discrimination in the F2 dimension is particularly vulnerable to hearing loss, whereas discrimination in the F1 dimension is not, is consistent with the relative importance of F1 in vowel contrasts in most languages. Additionally, the spectral magnitude of F2 is inherently lower than that of F2 (Klatt, 1980), exacerbating the effect of hearing loss in the F2 frequency region. We need to compare these thresholds for F2 discrimination in the presence of F1 to those for F2 formants without F1, presented at comparable sound levels (see Henry et al., 2017; Lyzenga and Horst, 1997, 1998).

ACKNOWLEDGEMENT
Supported by NIH-R01-DC001641

REFERENCES

Becker-Kristal, Roy (2010) Acoustic Typology of Vowel Inventories and Dispersion Theory: Insights from a Large Cross-Linguistic Corpus Ph.D. Dissertation, University of California, Los Angeles

Carney, LH (2018) Supra-threshold hearing and fluctuation contrast: Implications for sensorineural and hidden hearing loss, JARO 19:331-352. PMCID: PMC6081887

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), 1-12. e0004-15.2015 1–1. (DOI: 10.1523/ENEURO.0004-15.2015). PMCID: PMC4596011.

Carney, LH, Kim DO, 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,*. *Springer.* Advances in experimental medicine and biology. 894:427-35. PMCID:PMC5081182

Carney, LH, and JM McDonough (2019) Nonlinear auditory models yield new insights into representations of vowels, Atten Percept Psychophys, 81(4):1034-1046. DOI: 10.3758/s13414-018-01644-w. PMCID: PMC6581637.

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

Deng, L., & Geisler, C. D. (1987). A composite auditory model for processing speech sounds. *The Journal of the Acoustical Society of America*, *82*(6), 2001-2012.

Hamza, Y., Farhadi, A., Schwarz, D.M., McDonough, J.M., Carney, L.H. (2022, in review) Representations of fricatives in sub-cortical model responses: comparisons with human perception, BioRxiv.

Henry, K.S., Amburgey, K.N., Abrams, K.S., Idrobo, F., Carney, L.H. (2017) Formant-frequency discrimination of synthesized vowels in budgerigars (*Melopsittacus undulatus*) and humans, JASA. *142*(4), 2073-2083; doi: http://dx.doi.org/10.1121/1.5006912.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2395-2405.

Kim, D. O., Carney, L., & Kuwada, S. (2020). Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. *Journal of Neurophysiology*, *124*(4), 1198-1215.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*(3), 971-995.

Krishna, B. S., & Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *Journal of neurophysiology*, *84*(1), 255-273.

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*(1), 522-539.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological reviews*, *84*(2), 541-577.

Lindblom, B. (1986). Phonetic universals in vowel systems. In: Ohala, J., Jaeger, J. (eds.) Experimental Phonology. Orlando: Academic Press, 13-44.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. Language, 48, 839-862. doi:10.2307/411991

Lyzenga, J., & Horst, J. W. (1995). Frequency discrimination of bandlimited harmonic complexes related to vowel formants. *The Journal of the Acoustical Society of America*, *98*(4), 1943-1955.

Lyzenga, J., & Horst, J. W. (1997). Frequency discrimination of stylized synthetic vowels with a single formant. *The Journal of the Acoustical Society of America*, *102*(3), 1755-1767.

Lyzenga, J., & Horst, J. W. (1998). Frequency discrimination of stylized synthetic vowels with two formants. *The Journal of the Acoustical Society of America*, *104*(5), 2956-2966.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, *24*(2), 175-184.

Rosenberg, A. E. (1971). Effect of glottal pulse shape on the quality of natural vowels. *The Journal of the Acoustical Society of America*, *49*(2B), 583-590.

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. J. Phonetics 25:255-286

Tan and Carney (2006), Predictions of Formant-Frequency Discrimination in Noise Based on Model Auditory-Nerve Responses, J. Acoust. Soc. Am. 120:1435-1445. PMCID: PMC2572872.

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *66*(5), 1381-1403.

Zilany, M. S., & 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*(1), 402-417.

Zilany, M.S.A., Bruce, I.C., and L. H. Carney (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. J Acoust Soc Am 135:283-286. PMCID: PMC3985897.