

## **Neural Mechanisms for Timbre: Spectral-Centroid Discrimination based on a Model of Midbrain Neurons**

Braden N. Maxwell<sup>1,2,3</sup>, Johanna B. Fritzinger<sup>2</sup>, and Laurel H. Carney<sup>2,3†</sup>

<sup>1</sup>Department of Music Theory, Eastman School of Music, University of Rochester, Rochester, NY, USA

<sup>2</sup>Department of Neuroscience, University of Rochester, Rochester, NY, USA

<sup>3</sup>Department of Biomedical Engineering, University of Rochester, Rochester, NY, USA

† Corresponding author: [Laurel.Carney@Rochester.edu](mailto:Laurel.Carney@Rochester.edu)

### **Introduction**

The spectral centroid, or center of mass of the spectrum, has been shown to play a substantial role in the perception of timbre (McAdams, 2019); however, the neural code for the spectral centroid is not fully understood. Here we used computational models to test spectral-centroid encoding by auditory neurons. Amplitudes of slow fluctuations (slow temporal changes in firing rate) in auditory-nerve (AN) responses vary systematically across neurons tuned to frequencies near the spectral centroid. These changes in fluctuation amplitudes are reflected in the average discharge rates of midbrain neurons that are sensitive to fluctuations (Carney, 2018). We show that the spectral-centroid discrimination thresholds from Allen and Oxenham (2014) can be estimated based on population responses of model midbrain neurons (Carney & McDonough, 2019). Additionally, model midbrain representations of the spectral centroid are influenced by changes in the fundamental frequency (F0) of the stimulus, suggesting a sub-cortical basis for the interaction between pitch and the perception of timbre (Krumhansl & Iverson, 1992; Marozeau & de Cheveigné, 2007, Allen & Oxenham, 2014).

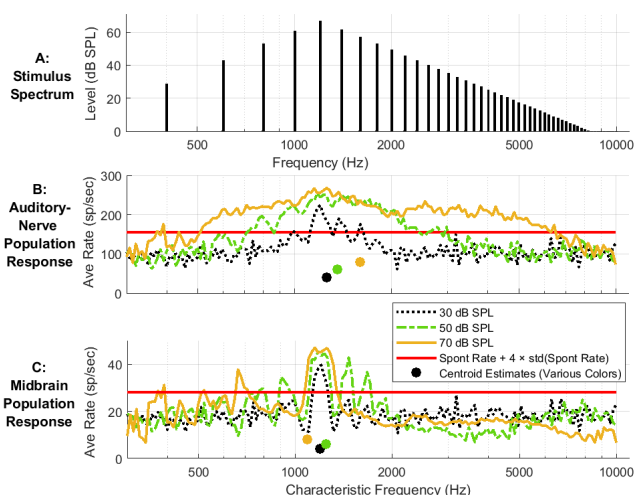
### **Method**

Spectral-centroid stimuli matched those of Allen and Oxenham (2014), experiments 1 and 2, including roving of stimulus parameters. Average rates of tonotopic populations at two stages of the auditory system in response to two-interval stimuli were simulated using computational models for the AN (Zilany, Bruce, & Carney, 2014) and midbrain (including cochlear nucleus stage; Carney & McDonough, 2019). Model populations included logarithmically spaced characteristic frequencies (CF, the frequency to which a neuron is most sensitive) over ranges specified below. Ten independent, high-spontaneous-rate, AN fibers and one midbrain neuron were simulated for each frequency channel. Midbrain simulations were based on band-suppressed neurons, a common midbrain cell type that is excited by unmodulated sounds and suppressed over a range of amplitude-modulation frequencies. For these simulations, the range of suppression was centered near 100 Hz, as commonly observed in the midbrain. The spectral centroid was generally reflected in increased activity of model neurons tuned near the peak of the spectral envelope (Fig. 1). Midbrain model responses were affected by the AN average rates and by the amplitudes of slow fluctuations in the AN responses.

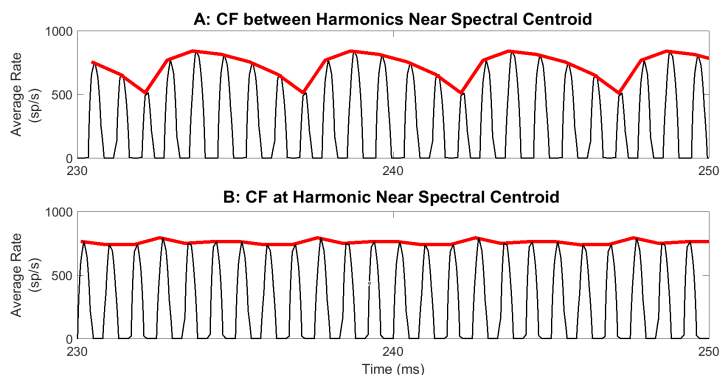
An estimate of the spectral centroid as represented in the model neural responses was based on the weighted average frequency of the population response. Weights were the average driven rates for each frequency channel, and  $\log(\text{CF})$  was used to calculate the weighted average. Driven rates were computed by subtracting an estimate of an upper bound on the spontaneous rate (mean + 4 × standard deviation of spontaneous rate; Fig. 1, red lines: 156 sp/s for AN, 28 sp/s for midbrain) from the average response rate for each frequency channel. This calculation was performed for model responses to the two intervals in each experimental trial, and the interval with the higher weighted average frequency was selected as the interval with the higher spectral centroid. The AN model included random internal noise; thus, the percent of correct model responses was computed based on 80 trials for each spectral centroid difference. A logistic function fit to the model %-correct results was used to estimate the difference limen (DL); a 70.7%-correct criterion for the DL was chosen to match the procedure in Allen and Oxenham (2014).

## Results

Figure 1 shows an example spectral-centroid stimulus (Allen & Oxenham, 2014) with AN and midbrain model population responses to stimuli at three sound levels. Centroid estimates based on each model response are shown in corresponding colors. Some aspects of the AN model are similar to other techniques for acquiring physiologically-based estimates of spectral centroid (e.g., Marozeau, de Cheveigné, McAdams, & Winsberg, 2003); for example, the AN model channels sample the stimulus spectrum on a logarithmic scale and filter widths change with frequency. However, the AN response used here was also shaped by several nonlinearities implemented in the Zilany et al. (2014) model. These nonlinearities realistically prevent the more densely packed components at high frequencies from resulting in higher average rates and cause the filter bandwidths (and response) to change with sound level (multi-colored curves, Fig. 1B), among other effects. The general flattening of the stimulus representation ultimately makes pitch interference more possible by emphasizing components farther from the spectral peak relative to the peak itself.



*Figure 2: Contrast in Fluctuations. (A) F0-fluctuations (red) in response of AN neuron with CF near centroid, but between harmonics (CF=1075 Hz). (B) Non-fluctuating (flat) response of AN neuron with CF at highest-magnitude harmonic (CF=1200 Hz). Stimulus is same as shown in Fig. 1A. Note: In Fig. 1C, band-suppressed midbrain neurons at CFs with fluctuating responses are suppressed, whereas those tuned to non-fluctuating channels have higher rate responses.*



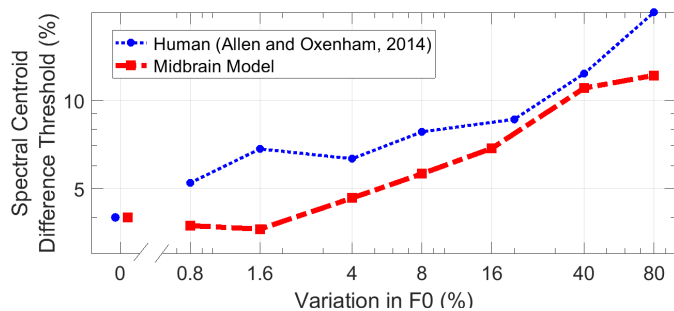
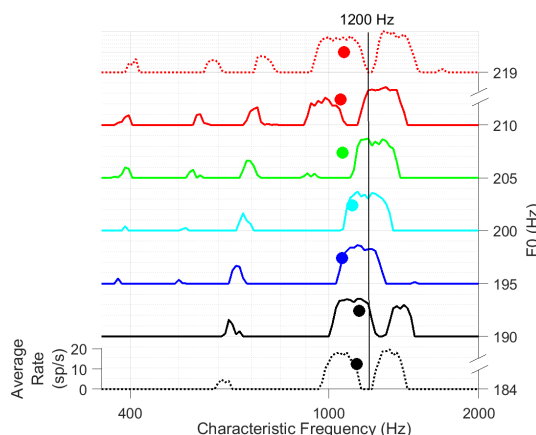
AN rates alone do not provide a suitable code for the spectral centroid. As shown in Fig. 1B, the sharp AN representation of the stimulus peak at a low sound level becomes broad at 70 dB SPL, a level in the range used for music listening (Epstein, Marozeau, & Cleveland, 2010) or conversational speech (Carney, 2018) and used in Allen and Oxenham (2014). This representation remains broad even if all spontaneous activity of the neurons is subtracted (average + 4 standard deviations, red line in Fig. 1B). The centroid estimate based on the AN response changes dramatically with increasing level, contrary to the general impression that the brightness of a sound is stable when the sound is rescaled over a wide range of levels. The midbrain population response offers a sharper, more precise spectral-centroid code that is consistent across a wide range of sound levels (Fig. 1C).

*Figure 1: Spectral-Centroid Encoding in Model Populations. (A) Example stimulus: spectral centroid at 1200 Hz, F0=200 Hz, 70 dB SPL (Allen and Oxenham, 2014). (B,C) Average rates across tonotopic auditory-nerve (AN) and midbrain populations, respectively, and centroid estimates (circles) for responses to three sound levels. Only response rates above thresholds based on spontaneous rate (horizontal red lines) were included in centroid estimates. Here 160 CFs were logarithmically spaced from 0.3-10 kHz, to illustrate the broad AN population response. Note: Due to spread of excitation, centroids based on model AN responses were more level-dependent than those based on midbrain responses.*

The enhanced code in the midbrain response reflects the influence of temporal patterns in AN responses. Figure 2 demonstrates these different temporal patterns, called neural fluctuations (Carney, 2018). Model AN fibers with CFs far from the centroid or *between* harmonics near the spectral centroid (Fig. 2A) have amplitude fluctuations at F0 due to beating between multiple harmonics. In contrast, AN CFs *at* harmonics near the spectral centroid (Fig. 2B) have minimal amplitude fluctuations at F0 because their response is dominated by a single harmonic, due to both peripheral filtering and to nonlinear ‘capture’ of the response (Carney, 2018). The band-suppressed midbrain model (Fig. 1C) is excited at CFs with minimal fluctuations (Fig. 2B) and inhibited at CFs with strong fluctuations (Fig. 2A). Thus, for this midbrain cell type, maximal responses occur for CFs at harmonics near the centroid, and the maxima are sharply delineated by midbrain suppression of responses at higher and lower frequencies. This sharp midbrain representation leads to a model threshold for centroid discrimination that is very close to that of humans (Fig. 4, leftmost points). Note that smaller peaks in the midbrain population response are often shifted away from harmonic frequencies (e.g., 900-Hz peak in Fig. 1C, green curve, corresponds to 1000-Hz harmonic). The response profile is strongly shaped by beating between harmonics, which is optimized when two components have equal magnitudes in a filter output. On the slopes of the spectral envelope, filters with equal magnitude components in their outputs are asymmetrically positioned between components (see Henry et al., 2017, for similar phenomenon).

The sharpened representation of harmonics in the midbrain response provides a mechanism for interaction between F0 and spectral centroid perception. Figure 3 shows driven responses for 1200-Hz centroids with varied F0s (increasing from bottom to top). As F0 increases over a moderate range, the major peak in the driven response shifts to higher frequencies, consistent with Allen and Oxenham (2014) experiment 3, in which spectral-centroid discrimination was easier for congruent trials (for which F0 moved in the same direction as the centroid) than for incongruent trials. However, peaks in response to lower harmonics pull the centroid estimate used here to lower frequencies, complicating the influence of F0 variation.

*Figure 3: Effect of Shifting F0 on Centroid Estimation. Peak of spectral envelope remained constant at 1200 Hz while F0 shifted (increasing, bottom to top). Driven model responses are shown (e.g. response minus an upper bound on spontaneous rate). Centroid estimates are indicated by circles. Double peaks for F0s of 219 and 184 Hz indicate response to harmonics numbered [5,6] and [6,7], respectively. 180 CFs were logarithmically spaced from 0.125-2 kHz.*



*Figure 4: Midbrain model thresholds (dashed red curve) for spectral-centroid discrimination with increasing random F0 variation, compared to human (musician) thresholds (replotted from Allen and Oxenham, 2014; dotted blue curve). Model population responses were for 186 CFs logarithmically spaced from 0.125-2.4 kHz. Simulation for 0% variation used slightly fewer CFs: 180, over the same range.*

Additional complications of F0 variation include: (1) large F0 variations may result in two harmonics that are equal in magnitude, and a shift in the ranks of harmonics near the centroid (dotted lines, Fig. 3); (2) the difference in the spectral centroid between intervals affects the prominence of different harmonics; (3)

between-trial shifts of F0 and spectral centroid for both intervals, called rove, further complicate the comparison. Nevertheless, when all of these factors from the original experiment were included, simulations of human thresholds using the midbrain population responses showed a pattern of F0 interference in spectral centroid discrimination that was generally similar to Allen and Oxenham's result (Fig. 4). More remains to be understood regarding the mechanisms driving this interference in the midbrain model.

## **Discussion**

These modeling results demonstrate a possible role in timbre perception for known response properties of sub-cortical auditory neurons. The basis for these model results is the sensitivity of the midbrain to slow fluctuations in AN responses. The plausibility of this mechanism is supported by the elevation in model thresholds for spectral centroid discrimination when the F0 was made variable across intervals (Fig. 4), as observed by Allen and Oxenham (2014). These model results are consistent with the hypothesis that the midbrain may play an important role in spectral-centroid encoding, and that pitch and timbre interactions may have origins in the early stages of the auditory system. Future tests of this model should include spectral centroids that are not aligned with the spectral peak and spectra with multiple spectral peaks. As long as one or more relatively high-magnitude harmonics influence the perceptual spectral centroid, the proposed neural code may pertain. It should be emphasized that the sharpness of the midbrain rate code results from sharp changes in temporal fluctuations in AN responses. Even if information is not extracted by the midbrain in the way described by these models, such timing information may be extracted by similar processes at other stages in the auditory pathway.

## **Acknowledgments**

This work was supported by NIH-NIDCD Grant No. DC010813.

## **References**

- Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *JASA*, *135*(3), 1371-1379.
- Carney, L. H., & McDonough, J. M. (2019). Nonlinear auditory models yield new insights into representations of vowels. *Attention, Perception, & Psychophysics*, *81*(4), 1034-1046.
- Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: Implications for sensorineural and hidden hearing loss. *Journal of the Association for Research in Otolaryngology*, *19*(4), 331-352.
- Epstein, M., Marozeau, J., & Cleveland, S. (2010). Listening habits of iPod users. *Journal of Speech, Language, and Hearing Research*, *53*, 1472-1477
- Henry, K. S., Abrams, K. S., Forst, J., Mender, M. J., Neilans, E. G., Idrobo, F., & Carney, L. H. (2017). Midbrain synchrony to envelope structure supports behavioral sensitivity to single-formant vowel-like sounds in noise. *Journal of the Association for Research in Otolaryngology*, *18*(1), 165-181.
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 739-751.
- Marozeau, J., & de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *JASA*, *121*(1), 383-387.
- Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *JASA*, *114*(5), 2946-2957.
- McAdams, S. (2019). The perceptual representation of timbre. In: Siedenburg K., Saitis C., McAdams S., Popper A., Fay R. (eds), *Timbre: Acoustics, Perception, and Cognition* (pp. 23-57). Springer Handbook of Auditory Research, vol 69. Springer, Cham.
- Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *JASA*, *135*(1), 283-286.