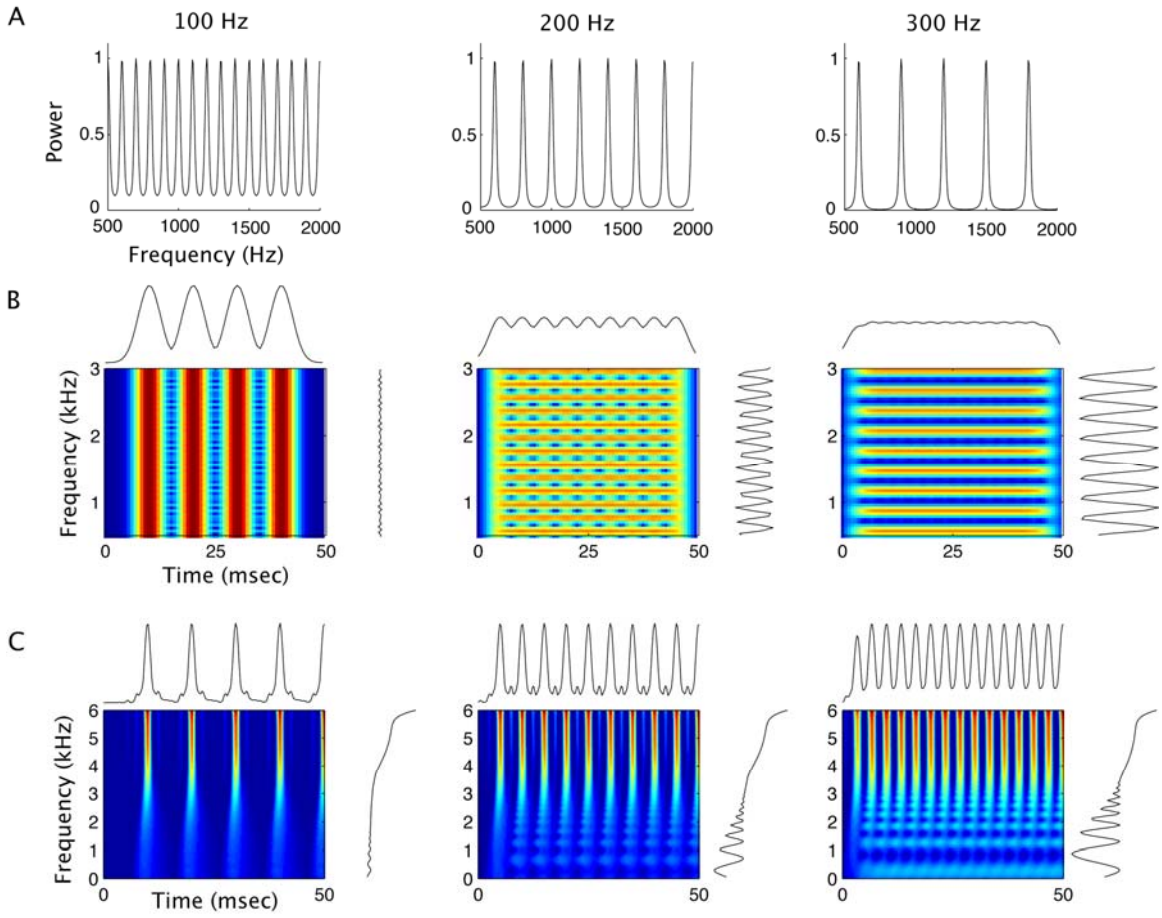
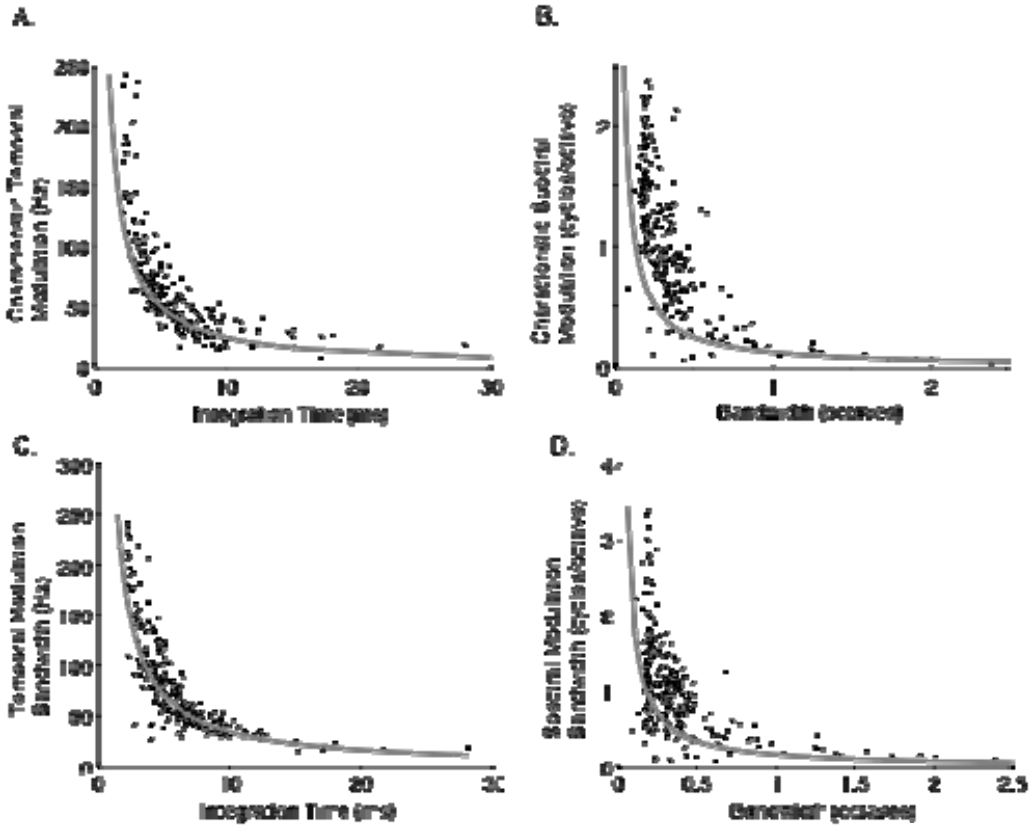
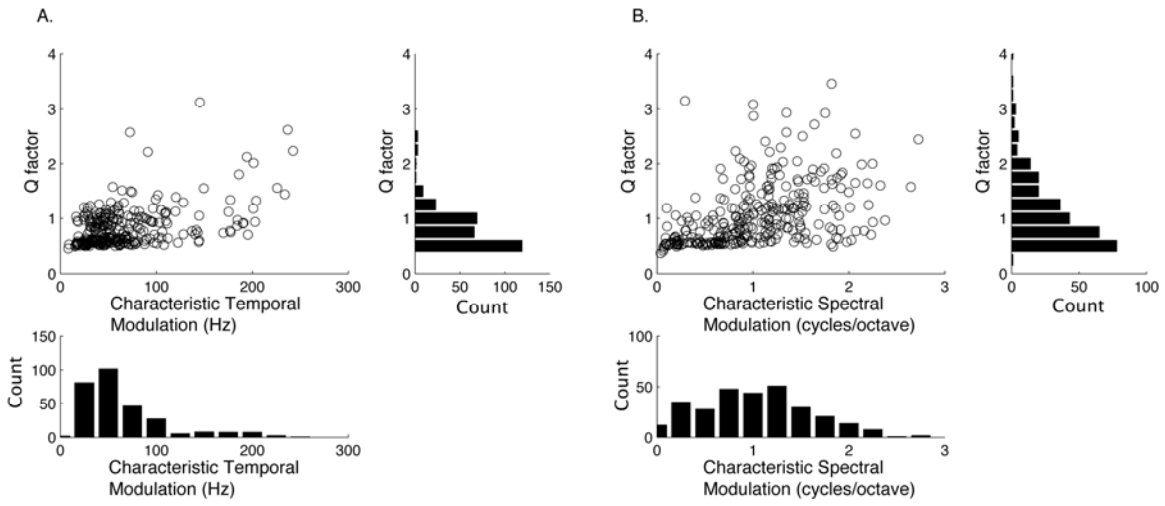



Figure S1: Spectrotemporal representation for harmonic complexes (cosine phase) through a constant bandwidth filterbank and the proposed peripheral auditory model (Gammatone filterbank with bandwidths that scal; see Methods). (A) Power spectrum of the harmonic complexes (100, 200 and 300 Hz, left to right; shown between 0.5 -2 kHz for clarity; harmonics extend to 20 kHz). (B) Spectrotemporal decomposition of the three harmonic complexes through a filtebank composed of constant bandwidth Gaussian filters (Gassian filters with 150 Hz bandwidth; Sighn & Theunissen 2003). Temporal and spectral marginals are shown above and to the right of each panel, respectively. For the 100 Hz complex the output is strictly temporal (no spectral modulation). This behavior occurs whenever two or more harmonics fall within each of the Gaussian filters (150 Hz bandwidth) and the harmonic complex is said to be unresolved (Schouten, 1940). By comparison, at 200 Hz there is a mixed spectrotemporal representation because the harmonics are partially resolved. For the 300 Hz harmonic complex, the harmonics are entirely resolved (1 harmonic per filter) and the filterbank produces strictly spectral modulations. (C) Peripheral auditory model decomposition of the three harmonic complexes enhances temporal modulations at the expense of spectral modulations. Temporal and spectral marginals are shown above and to the right of each panel, respectively. For the 100 Hz complex, the harmonics are entirely unresolved and the resulting spectrotemporal pattern consists strictly of temporal modulations. At 200 and 300 Hz there is a mixed representation. Low frequency channels are partially resolved so

that spectral modulations are visible at 200 and 300 Hz. By comparison, high frequency filters are unresolved and their outputs are predominantly temporal (for 100, 200 and 300 Hz). This frequency dependence results because filter bandwidths increase with increasing frequency in the model, as observed for auditory nerve fibers (Kiang 1965). Thus only low frequency channels have sufficient spectral resolution to resolve the harmonics and high frequency channels do a better job at preserving temporal modulations as observed for auditory nerve fibers (Jorris and Yin 1992). Despite this mixed representation the output pattern of the peripheral filterbank is predominantly temporal, as can be seen from the temporal marginal (*continuous line above spectrotemporal envelope*).




Supplementary Figure S2. Relationship between STRF preferences and modulation tuning. (A) Characteristic temporal modulation frequencies exhibit an inverse-like relationship with the STRF integration times. A similar trend is observed between characteristic spectral modulation frequency and STRF bandwidth (B). (C and D) STRF time-domain parameters (integration time and bandwidth) also exhibit an inverse-like dependence with spectral and temporal modulation bandwidths. Gray lines indicate the best power-law fits to the data. Although the Fourier transform and uncertainty principle require that integration time is inversely related to the system bandwidth (as in C and D), we point out that the observed inverse-like relationship between integration time and characteristic modulation (as in A and B) is not expected *a priori*.




Supplementary Figure S3: Sharpness of modulation tuning for the CNIC ensemble. Quality factors (Q, bandwidth to characteristic modulation frequency ratio) are plotted as function of cTMF and cSMF in A and B. The distributions for temporal and spectral Q-values are centered about ~1 (median quality factors Temporal: 0.7; Spectral: 0.89) with > 90% of neurons falling within Q=0.5 – 1.5 for temporal dimension and > 90% of neurons falling within Q=0.5 – 2 for spectral dimension. The median quality factor for temporal modulation closely match estimates for temporal modulation sensitivity in humans which have been estimated around 1 (Ewert and Dau, 2000). For both temporal and spectral modulations there is subtle dependence between Q-values and cTMF and cSMF (temporal: $r=0.44\pm0.07$, $p<0.01$; spectral: $r=0.41\pm0.07$, $p<0.01$) suggesting that neurons are more sharply tuned for high cTMF and cSMF.