Pitch of Harmonic Complex Tones: Rate Coding of Envelope Repetition Rate in the Auditory Midbrain

Yaqing Su[1,2] and **Bertrand Delgutte**[1,3]

[1] Eaton-Peabody Labs, Massachusetts Eye & Ear, Boston, MA, USA.

[2] Dept. of Biomedical Engineering, Boston University, Boston, MA, USA

[3] Dept. of Otolaryngology, Harvard Medical School, Boston, MA, USA

Summary

Envelope repetition rate (ERR) is an important cue for the pitch of harmonic complex tones (HCT), especially when the tone consists entirely of unresolved harmonics. Neural synchronization to the stimulus envelope provides a prominent cue for ERR in the auditory periphery, but this temporal code becomes degraded and gives way to rate codes in higher centers. The inferior colliculus (IC) likely plays a key role in this temporal-to-rate code transformation. Here we recorded single IC neuron responses to HCT at varying fundamental frequencies (F_0). ERR was manipulated by applying different inter-harmonic phase relationships. We identified a subset of neurons that showed a 'non-tonotopic' rate tuning to ERR between 160 and 1500 Hz. A comparison of neural responses to HCT and sinusoidally amplitude modulated (SAM) noise suggests that this tuning is dependent on the shape of stimulus envelope. A phenomenological model is able to reproduce the non-tonotopic tuning to ERR, and suggests it arises in the IC via synaptic inhibition.

1. Introduction

Harmonic complex tones present in speech, musical sounds and animal vocalizations evoke a strong pitch sensation at their fundamental frequency (F_0), even if they contain no energy at F_0 ("missing fundamental"). Although pitch plays important roles in speech and music perception and in the perceptual organization of auditory scenes, the neural mechanisms underlying pitch perception are still poorly understood. On the one hand, studies of the auditory nerve (AN) and cochlear nucleus (CN) have described multiple potential codes to pitch cues, including a rate-place code for resolved harmonics, temporal codes based on interspike interval distributions, and spatio-temporal codes dependent on both cochlear frequency selectivity and neural phase locking [1, 2, 3, 4]. On the other hand, pitch-selective neurons have been identified in a specific region of marmoset auditory cortex [5].

However, little is known about how the peripheral representations of pitch cues are integrated and transformed into a cortical pitch code. IC, the major nucleus in the auditory midbrain, is a logical target for addressing this question because it is the site of a transformation from a temporal code to a rate code for the frequency of amplitude modulations [6, 7].

Pitch percepts can arise either through the tonotopic pattern of low-numbered harmonics that are individually resolved by the cochlea or through neural phase locking to the envelope periodicity resulting from beating of unresolved harmonics encompassed by one auditory filter [8]. In general, the pitch produced by unresolved harmonics is less salient and more dependent on the inter-harmonic phase relationships than the pitch produced by resolved harmonics [8, 9].

Here, we investigate the neural representation of pitch cues by single-neuron recording from the IC in unanesthetized rabbits. Rabbits have good low frequency hearing like humans [10] and can discriminate F_0 of HCTs with missing fundamentals [11]. We characterize a rate code for the ERR that is unrelated to pure-tone frequency tuning and is likely derived from temporal cues in the auditory periphery. This "non-tonotopic" rate code could play a role in pitch perception based on unresolved harmonics.

2. Methods

2.1. Single-unit Recording

Extracellular spiking activity of 145 single IC neurons was recorded from 4 head-fixed, unanesthetized rabbits. All procedures were approved by the Animal Care Committee of Massachusetts Eye and Ear. Details of the preparation can be found in [12].

2.2. Stimuli

Acoustic stimuli were generated in MATLAB and diotically delivered by earphones coupled through custom- fitted ear molds. All stimuli were presented 10 times in random order, 200-ms on, 300-ms off.

To manipulate the shape and repetition rate of the temporal envelope, HCT consisting of equal-amplitude harmonics up to 18 kHz and with F_0 ranging from 26 to 2560 Hz were generated with three phase relationships among harmonics (Figure 1): 1) COS: all harmonics in cosine phase, ERR = F_0; 2) ALT: even harmonics in cosine phase and odd harmonics in sine phase, ERR = $2F_0$; 3) RAND: harmonics in random phase to minimize envelope fluctuations. The overall sound level was 40 dB SPL for all HCTs.

In 66 of the 145 neurons, SAM broadband noise (SAMN) was interleaved with HCTs at 40 dB SPL. Modulation frequencies (F_m) matched the F_0s of the HCTs. The modulation depth was always 1.

2.3. Data Analysis

For every neuron, we measured the average firing rate over the stimulus duration at each F_0 or F_m to create a rate- frequency profile (Figure 2). Spontaneous rate was averaged over the last 200 ms of the inter-stimulus interval. The neuron's best frequency (BF) was defined as the pure tone frequency that elicited the maximum firing rate at ~40dB SPL.

2.4. SFIE Model

We implemented the Same-Frequency Inhibition and Excitation (SFIE) model [13] to study possible neural mechanisms underlying the physiological results. This model is the cascade of a physiologically based AN model, a phenomenological ventral cochlear nucleus (VCN) cell model, and an IC cell model with the same structure as the VCN model (Figure 4a). VCN and IC models receive both excitatory post-synaptic potential (EPSP) and delayed inhibitory PSP (IPSP) in the form of an alpha function. Inhibition at the IC stage is both stronger and slower than at the VCN stage. Key model parameters and the range of values used in the simulation are shown in Table I.

3. Results

3.1. Non-tonotopic Rate Tuning to ERR

Figure 2a shows the firing rate of a neuron as a function of F_0 for an HCT in COS phase. The rate profile shows multiple peaks occurring when the F_0 is a small integer submultiple of the BF (BF, BF/2, BF/3, …), reflecting cochlear tuning to resolved harmonics. This rate-place code to resolved harmonics is inherited from the auditory nerve [1] and is not the focus of the present paper. In contrast, the rate-F_0 profile for the neuron in Figure 2b showed a single peak at 224 Hz for COS HCT. Though this frequency was near BF/2, there were no peaks at other submultiples of BF. This neuron gave a weak, nearly flat response to RAND HCT. Since RAND and COS stimuli have identical power spectra, the difference in response profile suggests that the neuron's tuning to COS HCT is unrelated to cochlear frequency selectivity but rather depends on temporal envelope fluctuations. With ALT HCT, the neuron exhibited bandpass tuning similar to COS, but shifted one octave towards lower F_0s. Since the ERR of ALT is twice the F_0, this neuron seems to be tuned to ERR, not F_0.

Among 145 neurons tested with HCT, 32% showed bandpass tuning to ERR as in Figure 2b. We defined the "best ERR" as the F_0 where a bandpass neuron reached its maximum firing rate in response to COS HCT. A "best ERR" was not defined for neurons with other tuning shapes to ERR (5% were band-reject, 6% lowpass, 14% highpass, 4% flat, 39% other).

Across the neuronal sample, best ERR for COS HCT spanned a wide frequency range, mainly 160–1500 Hz (Figure 3). The inset of Figure 3 shows a scatter plot of the best ERR vs. the pure tone BF for neurons tuned to ERR. There is no correlation between the two frequencies, justifying the term "non-tonotopic tuning" to ERR.

In 66 neurons, we compared rate tuning to ERR for HCT vs. SAMN. As with HCT, we defined the best ERR for SAMN as the F_m yielding the maximum firing rate. The neuron in Figure 2b showed bandpass tuning to the ERR of both COS and SAMN, but the best ERR was lower for SAMN. Among the 66 neurons tested with SAMN, only 10 showed bandpass tuning. While the best ERRs for SAMN were uncorrelated with BF (Figure 3 inset), they were restricted to a lower frequency range (112–448 Hz, Figure 3) than best ERRs for COS. For the 8 neurons showing bandpass tuning to both COS HCT and SAMN, the correlation between best ERR for the two stimuli was not statistically significant (r=0.51, p=0.2).

3.2. Model Response to HCT and SAM Stimuli

To explore possible mechanisms giving rise to the non- tonotopic tuning to ERR, we stimulated the SFIE model with HCT, SAMN and SAM tones (carrier frequency at the model's BF). Rate-frequency profiles are shown in Figure 4b-d for each model stage. At the AN stage, the model fired approximately 1 spike per ERR cycle for COS and ALT HCT up to 200 Hz, then the rate plateaued at higher frequencies. The rising portion of these profiles was maintained at subsequent stages and contributed to bandpass ERR tuning for COS and ALT at the IC stage. In contrast, AN firing rates for RAND and SAM were nearly constant over the entire frequency range. VCN responses were similar to AN except for a modest overshoot before the plateau with ALT and COS. Model IC responses resembled the neural data in Figure 2b: bandpass tuning to ERR for COS and ALT, and flat, weak response to RAND. Importantly, the model firing rate peaked one octave below for ALT compared to COS, consistent with the neural data. The model showed bandpass tuning for SAMN with a lower best ERR than for COS, consistent with the trends in the neural data.

We simulated the model with various parameter combinations (Table I). Rate tuning to ERR was observed in some model configurations, with best ERR for COS ranging from 48 to 190 Hz (not shown), which is lower than the range in the neural data. A critical factor for ERR tuning was the balance between inhibition and excitation at the IC stage. When inhibition was weaker than excitation, the IC output resembled the VCN response in Figure 4c, with minimal tuning. A decrease in firing rate at high frequencies, which is necessary for bandpass tuning, only occurred when the inhibition was stronger than excitation [13].

To understand the difference in tuning between SAMN and COS HCT observed in both the data and the model, the insets in Figure 4b and 4d show the temporal firing patterns of the AN stage and PSPs of the IC stage at 113 Hz, the model's best ERR for COS. The temporal firing patterns of the AN stage reflect the different stimulus envelope shapes-transient, impulse-like for COS, and more graded for SAMN. As explained by Nelson and Carney [13], the PSPs of the VCN and IC stages turn from phasic to tonic with increasing input frequency. With COS stimuli at 113 Hz, the transient EPSPs at the input to the IC stage occur out of phase with the nearly sinusoidal IPSP, allowing strong firing. In contrast, the EPSPs for SAMN have lower amplitude, and the IPSP is sustained, resulting in weak firing. Thus, the dynamics of excitation and inhibition in the model interact with the different envelope shapes of COS and SAMN to yield a lower best ERR for SAMN.

4.   Discussion

4.1.   Implication for Pitch Processing

We identified a non-tonotopic rate code for ERR of HCT in the IC of unanesthetized rabbits. This finding extends the results in [14], where multi-unit clusters in the IC of anesthetized guinea pigs showed bandpass or band-reject rate tuning to ERR of HCT in sine and ALT phase. Best ERRs in their study ranged from 50 to 400 Hz, with a mode at 141 Hz. This is clearly lower than the 160–1500 Hz range we observed, in part because they only tested F_0 up to 400 Hz, and perhaps also because they used an anesthetized preparation.

Tuning to ERR could be reproduced by the SFIE model, thereby suggesting a mechanism for how the peripheral temporal code for ERR is transformed into a rate code in the midbrain. However, ERR and pitch are not equivalent. For example, the pitch of an ALT HCT containing resolved harmonics is matched to its F_0, not ERR [9]. Nevertheless, the non-tonotopic rate code for ERR could play a role in extracting the pitch of unresolved harmonics. This information could be integrated with the more reliable pitch information provided by resolved harmonics at a central "pitch center". The non-tonotopic rate code may be especially important in listeners with sensorineural hearing loss, where reduced cochlear frequency selectivity results in a degraded representation of resolved harmonics [15].

4.2. ERR Tuning Depends on Envelope Shape

The differences in ERR tuning between HCT and SAMN suggests that the tuning is dependent on envelope shape, consistent with previous studies in anesthetized cat [16], chinchilla [17] and gerbil [18, 19] that also found IC neurons respond differently to AM stimuli with various envelope shapes. The non-tonotopic rate code to ERR in IC, therefore, is confounded by the sensitivity to envelope shape.

4.3. Limitations of SFIE Model

The SFIE model was able to simulate non-tonotopic rate tuning over a range of best ERRs for COS HCT. However, the range of best ERR that could be generated in the model was restricted to lower frequencies (48–190 Hz) compared to the neural data (160–1500 Hz). Because the ERR tuning is dependent on the transient, synchronized response of AN and CN stages, adding multiple input pathways to IC might enhance the strong fluctuations of PSPs at high frequencies, thereby extending the range of best ERR. As expected, the model only simulated bandpass tuning; band-reject and other complex response patterns were not reproduced. More recent elaborations of the SFIE model with more complex structures can account for other tuning types with SAM stimuli [20], although they have not yet been systematically tested with HCT.

References

[1]. Cedolin L, Delgutte B: Pitch of complex tones: rate- place and interspike interval representations in the auditory nerve. J. Neurophysiol 94 (2005) 347–362. [PubMed: 15788522]

[2]. Cedolin L, Delgutte B: Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. J. Neurosci 30 (2010) 12712–12724. [PubMed: 20861376]

[3]. Cariani PA, Delgutte B: Neural correlates of the pitch of complex tones. I: Pitch and pitch salience. J. Neurophysiol 76 (1996) 1698–1716. [PubMed: 8890286]

[4]. Winter IM, Palmer AR, Wiegrebe L, Patterson RD: Temporal coding of the pitch of complex sounds by presumed multipolar cells in the ventral cochlear nucleus. Speech Comm. 41 (2003) 135–149.

[5]. Bendor D, Wang X: The neuronal representation of pitch in primate auditory cortex. Nature 436 (2005) 1161. [PubMed: 16121182]

[6]. Joris PX, Schreiner CE, Rees A: Neural processing of AM sounds. Physiol. Rev 84 (2004) 541–577. [PubMed: 15044682]

[7]. Nelson PC, Carney LH: Neural rate and timing cues for detection and discrimination of AM tones in the awake rabbit inferior colliculus. J. Neurophysiol 97 (2007) 522–539. [PubMed: 17079342]

[8]. Plack CJ, Oxenham AJ: The psychophysics of pitch - In: Pitch. Springer, New York, 2005, 7–55.

[9]. Shackleton TM, Carlyon RP: The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J. Acoust Soc. Am 95 (1994) 3529–3540. [PubMed: 8046144]

[10]. Heffner HE, Heffner RS: Hearing ranges of laboratory animals. J. Am. Assoc. Lab. Anim Sci 46 (2007) 20–22.

[11]. Delgutte B, Gelman A, Chungm Y: Rabbits can discriminate harmonic complexes with missing fundamentals. ARO Abstr. 42 (2018) #797.

[12]. Zuk N, Delgutte B: Neural coding of time-varying interaural time differences and time-varying amplitude in the inferior colliculus. J. Neurophysiol 118 (2017) 544–563. [PubMed: 28381487]

[13]. Nelson PC, Carney LH: A phenomenological model of peripheral and central neural responses to AM tones. J. Acoust Soc. Am 116 (2004) 2173–2186. [PubMed: 15532650]

[14]. Shackleton TM, Liu LF, Palmer AR: Responses to diotic, dichotic, and alternating phase harmonic stimuli in the inferior colliculus of guinea pigs. JARO 10 (2009) 7690.

[15]. Moore BC: The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. JARO 9 (2008) 399–406. [PubMed: 18855069]

[16]. Zheng Y, Escabi MA: Distinct roles for onset and sustained activity in the neuronal code for temporal periodicity and acoustic envelope shape. J. Neurosci 28 (2008) 14230–14244. [PubMed: 19109505]

[17]. Sinex DG, Henderson J, Li H, Chen GD: Responses of chinchilla inferior colliculus neurons to AM tones with different envelopes. JARO 3 (2002) 390–402. [PubMed: 12486595]

[18]. Krebs B, Lesica NA, Grothe B: The representation of amplitude modulations in the mammalian auditory midbrain. J. Neurophysiol 100 (2008) 1602–1609. [PubMed: 18614754]

[19]. Schnupp JW, Garcia-Lazaro JA, Lesica NA: Peri- odotopy in the gerbil inferior colliculus: local clustering rather than a gradient map. Frontiers Neur. Circ 9 (2015) 37.

[20]. Carney LH, Li T, McDonough JM: Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. Eneuro 2 (2015) e0004–15.
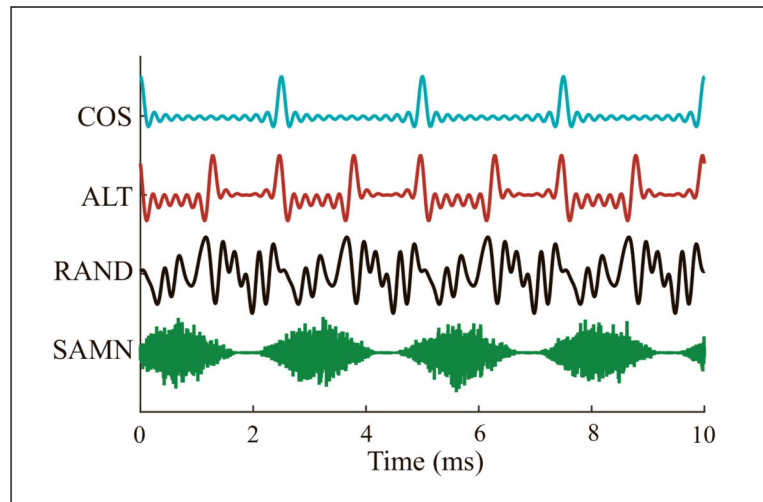
Figure 1.
Temporal waveforms of HCT with three inter-phase relationships at $F_0 = 400$ Hz, and SAM noise at $F_m = 400$ Hz. COS: ERR = F_0, ALT: ERR = $2F_0$, RAND: flat envelope, SAMN: ERR = F_m, and envelope shape different from HCT.
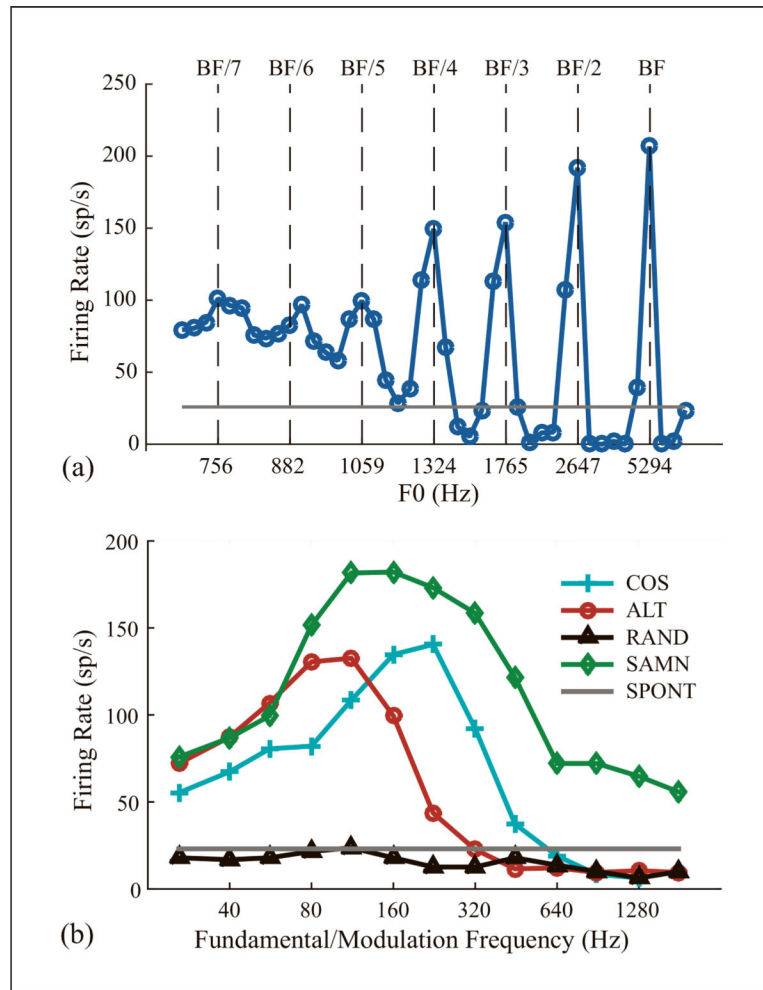
Figure 2.
Firing rate profiles of two neurons tuned to different attributes of the stimuli. (a) Rate profile of the neuron showed local maxima at BF/n, n = 1, 2, ..., 5. (b) Neuron (BF = 500 Hz) showed BP tuning to ERR of COS and ALT HCT, and SAMN. Best ERR is 190 Hz for COS, 129 Hz for SAMN.
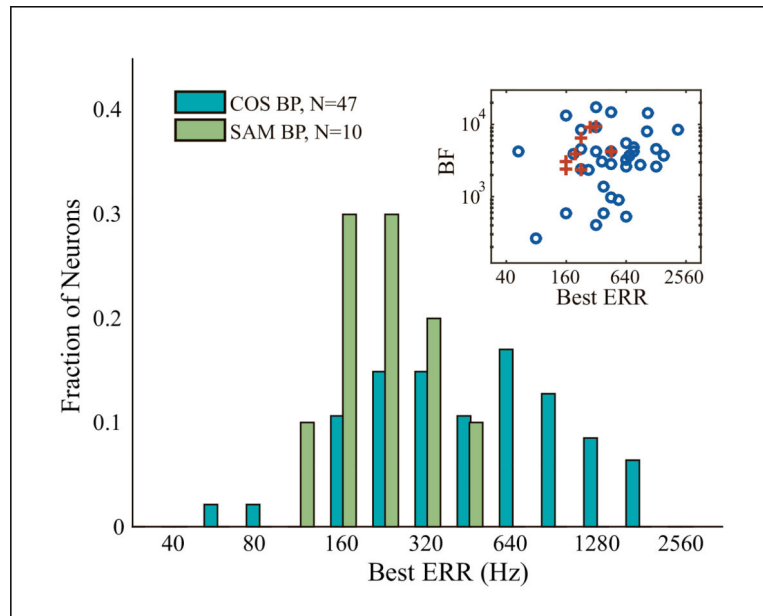
Figure 3.
Distribution of best ERR for COS HCT and SAMN. Inset: individual neuronSs best ERR for COS (N = 35, circle) and SAMN (N = 8, cross) showed no correlation with BF. The number of neurons included is smaller than in the histogram because BF could not be identified in some neurons with complex response pattern to pure tones.
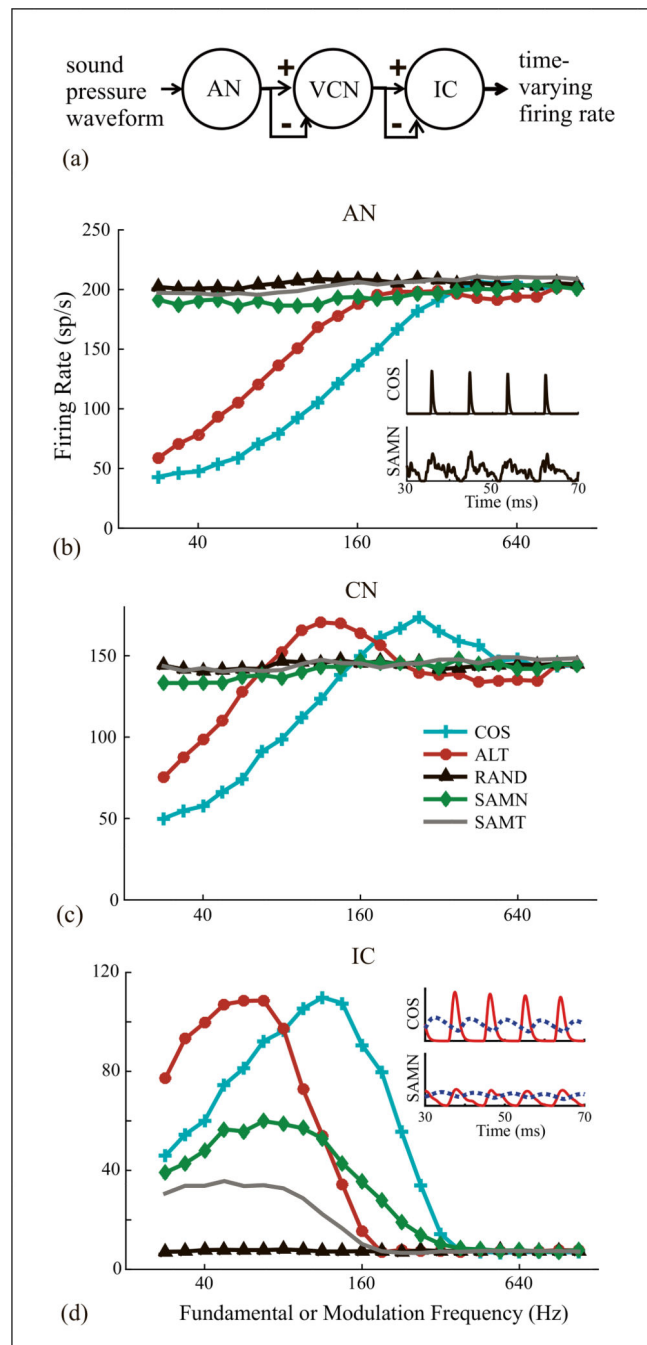
Figure 4.
Simulated rate-frequency profiles of SFIE model in response to HCT and SAM stimuli. (a) Model diagram. (b) AN model output. Inset: steady state firing pattern at ERR = 113 Hz for COS HCT and SAMN (SAMN y axis zoomed in 4x). (c) CN model output. (d) IC model output. Inset: steady state EPSP (continuous) and IPSP (dashed) at ERR = 113 Hz. Parameters: BF = 12.8kHz, CN inh_str = 0.6, t_{inh} = 2ms, te_x = 0.5 ms, IC inh_str = 1.5, t_{inh} = 3 ms, t_{ex} = 0.5 ms.

Table I.

Key VCN/IC model parameters.

	VCN	IC
Inhibition delay re. excitation (ms)	1–3	1–8
EPSP time constant τ_{ex} (ms)	0.3–0.5	0.5–2
IPSP time constant τ_{inh} (ms)	0.5–2	0.5–10
inhibition strength re. excitation inh_str	0–1.8	0–1.8