Behavioral/Systems/Cognitive

Ambiguous Pitch and the Temporal Representation of Inharmonic Iterated Rippled Noise in the Ventral Cochlear Nucleus

Mark Sayles and Ian M. Winter

Centre for the Neural Basis of Hearing, The Physiological Laboratory, Cambridge CB2 3EG, United Kingdom

Neural coding of the pitch of complex sounds is vital for animals' ability to communicate and to perceptually organize natural acoustic scenes. Harmonic complex sounds typically have a well defined pitch corresponding to their fundamental frequency, whereas inharmonic sounds can exhibit pitch ambiguity: their pitch can have more than one value. Iterated rippled noise (IRN), a common "pitch stimulus," is generated from broadband noise by a cascade of delay-and-add steps, with the delayed noise phase-shifted by ϕ degrees. By varying ϕ, the (in)harmonicity, and therefore the pitch ambiguity, of IRN can be manipulated. Recordings were made from single-units in the ventral cochlear nucleus of anesthetized guinea pigs in response to IRN and complex tones, systematically varying the inharmonicity. In their all-order interspike interval distributions, primary-like and chopper units tuned within the phase-locking range of best frequencies represent the waveform temporal fine structure (which varies with ϕ). In contrast, those units tuned to higher frequencies represent the temporal-envelope modulation (independent of ϕ). We show a temporal representation of ambiguous pitch for IRN and complex tones based on responses to the stimulus fine structure. Within the dominance region for pitch this representation follows the predictions of classic human behavioral experiments and provides a unifying contribution to possible neuro-temporal explanations for the pitch shift and pitch ambiguity associated with many inharmonic sounds.

Key words: pitch ambiguity; complex tone; inharmonic sounds; iterated rippled noise; cochlear nucleus; interspike intervals; auditory brainstem

Introduction

Vocal communication, musical melody recognition, and the perceptual organization of sounds into meaningful "auditory objects" rely on accurate neural coding of pitch (Bregman, 1990; Plack et al., 2005). Whether the brain uses temporal (Schouten, 1940; Licklider, 1951) or spectral (Goldstein, 1973; Wightman, 1973a; Terhardt, 1974) information to determine a sound's pitch remains an open question (Cedolin and Delgutte, 2005; de Cheveigné, 2005); however, recent evidence favors temporal processing with some dependence on spectral place (Oxenham et al., 2004; Bernstein and Oxenham, 2005). Temporal information reaches the central auditory system by auditory nerve fibers (ANFs) "phase-locking" to basilar-membrane vibrations; the product of a rapid fine-structure, and a slower temporal-envelope vibration. The fundamental frequency (F0) of harmonic complex sounds may be extracted by neural processing of either one, or both of these temporal cues. Applying a frequency

shift to each harmonic of a complex sound, making the spectrum inharmonic, alters the fine structure but not the envelope. Inharmonic sounds often evoke more than one pitch (pitch ambiguity), none of which corresponds to the envelope periodicity (pitch shift), and therefore provide evidence for fine-structure based pitch (Schouten, 1940; de Boer, 1956b; Schouten et al., 1962).

Neurophysiological studies have examined the processing of harmonic and inharmonic complex sounds from the auditory nerve (Javel, 1980; Palmer and Winter, 1993; Simmons and Ferragamo, 1993; Rhode, 1995; Cariani and Delgutte, 1996a,b; Cedolin and Delgutte, 2005, 2007) to the auditory cortex (Bendor and Wang, 2005). However, studies of inharmonic sounds are limited to narrowband amplitude-modulated (AM) tones (Javel, 1980; Simmons and Ferragamo, 1993; Rhode, 1995; Cariani and Delgutte, 1996a). Broadband signals known as iterated rippled noises (IRNs) are popular pitch stimuli (Patterson et al., 2002; Bendor and Wang, 2005; Schönwiesner and Zatorre, 2008), and fine-structure detection in low-frequency "dominance" regions is thought to be the basis of IRN pitch (Bilsen and Ritsma, 1967/68, 1969/70; Bilsen, 1966; Yost and Hill, 1979; Yost, 1996; Yost et al., 1996; Shofner and Yost, 1997). The unambiguous pitch of harmonic IRN is represented in the firing patterns of ANFs (Fay et al., 1983; ten Kate and van Bekkum, 1988) and cochlear-nucleus neurons (Bilsen et al., 1975; Shofner, 1991, 1999; Winter et al., 2001; Sayles and Winter, 2007) by action potentials locked to either the fine-structure or envelope periodicity. The pitch of

inharmonic IRN can be ambiguous (Bilsen, 1966; Raatgever and Bilsen, 1992; Yost, 1997), and the (in)harmonicity (and pitch ambiguity) can be manipulated by varying a single parameter.

Here we examine the temporal representation of the pitch-shift and ambiguity of IRN in the responses of anesthetized guinea-pig ventral cochlear nucleus (VCN) units. In their interspike-interval distributions, VCN primary-like units and chopper units tuned to the dominance region represent the pitch-shift and ambiguity of IRN and inharmonic complex tones predicted by theoretical and psychophysical studies (Fourcin, 1965; Bilsen, 1966; Yost and Hill, 1979; Yost, 1996, 1997). The aim of the present work is to provide neurophysiological evidence for fine-structure detection by VCN units underlying the pitch-shift and ambiguity of broadband inharmonic signals.

This work has been presented in abstract form at "Acoustics '08", Paris, France (Sayles and Winter, 2008a).

Materials and Methods

The preparation. Experiments were performed on 20 pigmented guinea pigs (*Cavia porcellus*), weighing between 320 and 600 g. Animals were anesthetized with urethane (1.0 g/kg, i.p.). Hypnorm (fentanyl citrate, 0.315 mg/ml; fluanisone, 10 mg/ml; Janssen) was administered as supplementary analgesia (1 ml/kg, i.m.). Anesthesia and analgesia were maintained at a depth sufficient to abolish the pedal withdrawal reflex (front paw). Additional doses of hypnorm (1 ml/kg, i.m.) or urethane (0.5 g/kg, i.p.) were administered on indication. Core temperature was monitored with a rectal probe and maintained at 38°C using a thermostatically controlled heating blanket (Harvard Apparatus). The trachea was cannulated and on signs of suppressed respiration, the animal was ventilated with a pump (Bioscience). Surgical preparation and recordings took place in a sound-attenuated chamber (IAC). The animal was placed in a stereotaxic frame, which had ear bars coupled to hollow specula designed for the guinea-pig ear. A mid-sagittal scalp incision was made and the periosteum and the muscles attached to the temporal and occipital bones were removed. The bone overlying the left bulla was fenestrated and a silver-coated wire was inserted into the bulla to contact the round window of the cochlea for monitoring compound action potentials (CAP). The hole was resealed with Vaseline. The CAP threshold was determined at selected frequencies at the start of the experiment and thereafter after indication. If thresholds deteriorated by >10 dB and were nonrecoverable (e.g., by removing fluid from the bulla or by artificially ventilating the animal) the experiment was terminated. A craniotomy was performed exposing the left cerebellum. The overlying dura was removed and the exposed cerebellum was partially aspirated to reveal the underlying cochlear nucleus. The hole left from the aspiration was then filled with 1.5% agar in saline to prevent desiccation. The experiments performed in this study have been performed under the terms and conditions of the project license issued by the United Kingdom Home Office to the second author.

Neural recordings. Single units were recorded extracellularly with glass-coated tungsten microelectrodes (Merrill and Ainsworth, 1972). Electrodes were advanced in the sagittal plane by a hydraulic microdrive (650 W; David Kopf Instruments) at an angle of 45°. Single units were isolated using broadband noise as a search stimulus. All stimuli were digitally synthesized in real-time with a PC equipped with a DIGI 9636 PCI card that was connected optically to an AD/DA converter (ADI-8 DS; RME audio products). The AD/DA converter was used for digital-to-analog conversion of the stimuli as well as for analog-to-digital conversion of the amplified (×1000) neural activity. The sample rate was 96 kHz. The AD/DA converter was driven using ASIO (Audio Streaming Input Output) and SDK (Software Developer Kit) from Steinberg (Lloyd, 2002).

After digital-to-analog conversion, the stimuli were equalized (phonic graphic equalizer, model EQ 3600; Apple Sound) to compensate for the speaker and coupler frequency response and fed into a power amplifier (Rotel RB971) and a programmable end attenuator (0–75 dB in 5 dB steps, custom build) before being presented over a speaker (Radio Shack

30-1777 tweeter assembled by Mike Ravicz, Massachusetts Institute of Technology, Cambridge, MA) mounted in the coupler designed for the ear of a guinea pig. The stimuli were monitored acoustically using a condenser microphone (Brüel & Kjær 4134) attached to a calibrated 1 mm diameter probe tube that was inserted into the speculum close to the eardrum. Neural spikes were discriminated in software, stored as spike times on a PC and analyzed off-line using custom-written Matlab programs (The MathWorks).

Unit classification. After isolation of a unit, its best frequency (BF) and excitatory threshold were determined using audio-visual criteria. Spontaneous activity was measured over a 10-second period. Single units were classified based on their peri-stimulus time histograms (PSTH), the first-order interspike-interval distribution and the coefficient of variation (CV) of the discharge regularity. The CV was calculated by averaging the ratios of the mean ISI (interspike interval) and its SD between 12 and 20 ms after onset (Young et al., 1988). PSTHs were generated from spike-times collected in response to 250 sweeps of a 50 ms tone at the unit's BF at 20 and 50 dB above threshold. Tones had 1 ms $\sin^2$-on and $\cos^2$-off gates, their starting phase was randomized, and they were repeated with a 250 ms period. PSTHs were classified as primary-like (PL), primary-like with a notch (PN), chopper-sustained (CS), chopper-transient (CT), and onset-chopper (OC). For some units with very low BFs (<~0.5 kHz) it was not possible to assign them to one of the above categories. In the absence of a definitive classification these are grouped together as "low frequency" (LF) units.

Complex stimuli. IRN was generated from a noise waveform with a Gaussian distribution of instantaneous amplitudes. This waveform is delayed by time d, phase shifted by ϕ degrees (independent of frequency), and added back to the input waveform. This process is repeated for n iterations. The phase shift was implemented in the frequency domain, with ϕ varied in 30° steps between 0 and 330°. Delay d was varied in octave steps between 2 and 16 ms, corresponding to F0 s between 500 and 62.5 Hz. For some units, values of d were chosen to place harmonics 1–10 of the IRN signal at unit BF when $\phi = 0°$. Stimuli were generated with 16 iterations of the delay-and-add circuit, with the output of each iteration step serving as the input to the next ["add same," (Yost, 1996)]. IRN signals were low-pass filtered in the frequency domain at 10 kHz. In contrast to previous studies which described IRN signals with the parameters d, gain g, and n, we describe the IRN stimuli used in this study as IRN$[d, \phi, n]$ throughout.

Complex tones were generated in the time domain from a sum of sinusoids added either in cosine (COS) or random (RAND) phase. Each complex contained all (equal-amplitude) components of a series between 0 and 10 kHz. The frequency spacing (f) between components was varied in octave steps between 500 and 62.5 Hz. The frequency shift Δf applied to each component, moving it upward in frequency from the harmonic condition, in which components are all integer multiples of f, was calculated as $\Delta f = \phi f/360$, with ϕ varied in 30° steps between 0 and 330°, as for IRN signals. When $\Delta f > 0$ the lowest frequency component present in the physical stimulus was at Δf. The complex tone stimuli are described as COS$[d, \phi]$ and RAND$[d, \phi]$ throughout, to facilitate comparison with the equivalent IRN$[d, \phi, n]$ signals.

We presented a control stimulus of low-pass filtered (at 10 kHz) Gaussian noise (GN) at the same level as the pitch stimuli. All stimuli (IRN, COS, RAND, and GN) were 0.5 s in duration, presented with a 1 s repetition period, gated with 5 ms $\sin^2$-on and $\cos^2$-off ramps, and were part of a single array presented in an interleaved manner in random order for 25 repetitions. Each stimulus repetition was generated in real time from a new noise waveform (IRN and GN), or with a new random set of starting phases (RAND complex tones). Before the presentation of the complex stimuli, we collected a rate-level function in response to GN of 0.5 s duration. The complex stimuli were then presented at a sound level corresponding to the ~50% point on the noise rate-level function. Across all units, this corresponded to an overall sound level of between 27 and 65 dB sound pressure level (SPL) (mean, 44.5 dB SPL). Figure 1 shows example magnitude spectra, waveform autocorrelation functions (ACFs), and Hilbert-envelope ACFs for IRN signals with $d = 1$ ms (F0 = 1 kHz). The spectral representations (Fig. 1 *A*, *B*) demonstrate that with increasing ϕ, the stimulus spectrum drifts upward along the frequency

Magnitude Spectra

Waveform ACF

Envelope ACF
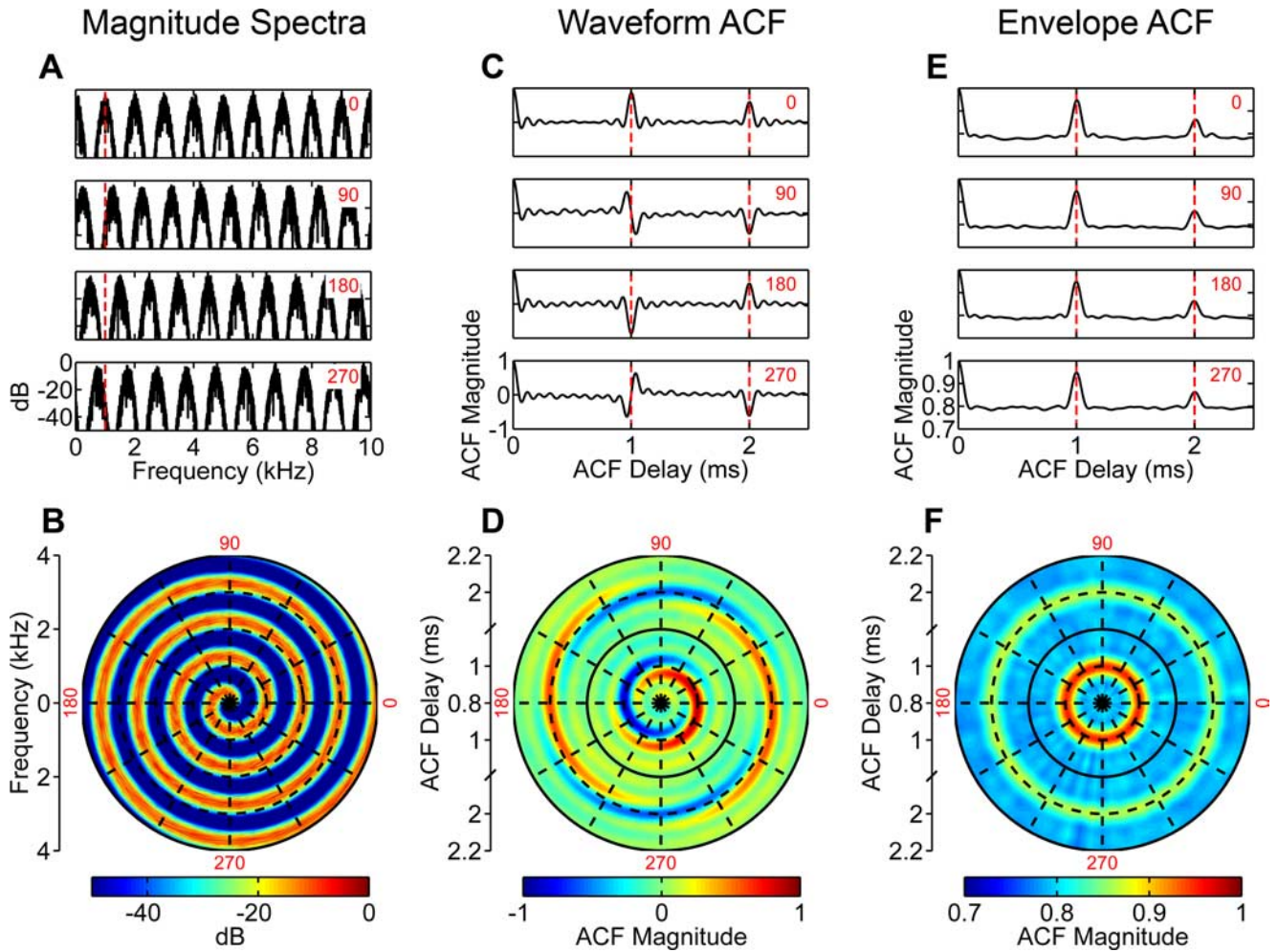
Figure 1. Frequency- and time-domain representations of IRN signals as a function of ϕ. **A**, Magnitude spectrum of IRN[1, ϕ, 16] for $\phi = 0$, 90, 180 and 270°, expressed as dB relative to the maximum. We applied a Hanning window and zero-padded the waveforms to 2^{16} points before computing the fast Fourier transform. The phase shift ϕ in each condition is indicated in the top right of each panel (red). **B**, Magnitude spectrum of the same signals as in **A**, plotted on polar coordinates as a function of ϕ. Circular dashed black lines are at N/d Hz, where N is an integer. Radial dashed black lines are at 30° intervals along the circumferential axis. The parameter ϕ is indicated around the circumference of the plot in red. For clarity the spectrum is shown between 0 and 4 kHz, although the signal bandwidth is 10 kHz. **C**, Waveform ACFs for the same signals as in **A**. **D**, ACFs plotted on polar coordinates, similar to the spectra in **B**. The radial axis extends from 0.8–1.2 ms, then, after a break in the axis, from 1.8–2.2 ms. The circular dashed lines are at d and $2d$ ms. The solid black line represents the axis break between 1.2 and 1.8 ms. **E**, Hilbert envelope ACFs as a function of ϕ for the same signals as in **A** and **B**. **F**, Same as **D** but for Hilbert envelope ACFs. **A–F**, As in the experiments, signals were 0.5 s long, sampled at 96 kHz, and low-pass filtered at 10 kHz.

axis. When $\phi = 0°$, there are peaks in the spectrum at integer multiples of $1/d$ Hz, by $\phi = 180°$ the spectral peaks are at odd-integer multiples of $1/2d$ Hz, and by $\phi = 360°$ the spectrum has returned to the "harmonic" condition. The data plotted on polar coordinates and interpolated in Figure 1B demonstrate that varying the parameter ϕ (along the circumferential axis) results in a continuous shift in the spectrum along the (radial) frequency axis. For display purposes, the spectrum is only shown between 0 and 4 kHz. The dashed black lines in Figure 1B are at integer multiples of $1/d$ Hz, corresponding to the position of the harmonic spectral peaks (red) when $\phi = 0°$, and corresponding to the spectral troughs (blue) when the signal is perfectly inharmonic ($\phi = 180°$). The waveform ACFs (Fig. 1C,D) show peaks (red) at d and $2d$ ms when $\phi = 0°$, a null (blue) at d and a peak at $2d$ when $\phi = 180°$, a transition from a peak to a null centered at d with a null at $2d$ when $\phi = 90°$, and a transition from a null to a peak centered at d with a null at $2d$ when $\phi = 270°$ (Fig. 1C,D). These differences in the waveform ACF reflect the effect of ϕ on the temporal fine-structure. In contrast, for all values of ϕ the Hilbert envelope ACF shows peaks at all integer multiples of d (Fig. 1E,F). Therefore, if a neuron responds to the temporal fine structure, its response will be modulated by changes in ϕ, whereas if it responds to the temporal envelope modulation, the response will be independent of changes in ϕ.

Analysis. We calculated the all-order interspike-interval distribution for each unit's response to each stimulus condition and constructed

interspike-interval histograms (ISIHs) with 50-μs wide bins between 0 and $2.5d$. ISIHs are expressed as firing rate by dividing raw bin-counts by the product of bin-width and the total number of spikes (Abeles, 1982; Shofner, 1999) and then smoothed with a sliding 0.45 ms Hanning-window. To determine which interspike intervals (if any) occur significantly more often in response to the complex pitch-evoking stimuli than in response to GN, we performed a Fisher-Pitman permutation test (Berry et al., 2002) on the interspike-interval distributions with the null hypothesis that at each bin the difference between the signal and noise responses is zero. First the difference between the signal and noise response (the observed interval-difference function) is calculated. Next, the interspike-interval distributions from the noise and the signal are pooled and re-sampled without replacement to give two populations of interspike intervals corresponding in size to the signal and noise interval distributions: the permutation distributions. The observed data are resampled in this way for a large number of replications (here, 5000). On each replication, the data are binned in the same way as for the observed data, and the difference between the two populations is calculated. The p value associated with each bin in the observed interval-difference function is then the proportion of replications on which the difference in the permutation distribution exceeds that in the observed data. We considered bins in the interval difference histogram with $p < 0.02$ to be significant.

Results

We recorded responses to harmonic and inharmonic IRN and complex tones from 84 isolated single units (33 PL, 6 PN, 25 CT, 2 CS, 11 LF, and 7 OC) in the VCN of 20 urethane-anesthetized guinea pigs. The majority of units included in this study had BFs in the range of phase locking. For the PL/PN group, 36 had BFs between 0.5 and 3 kHz, and three had BFs between 5 and 6.25 kHz. The CT/CS group had BFs between 0.5 and 5 kHz, with the majority of units (15 of 27) having BFs in the phase-locking range <1.25 kHz. Low frequency units had BFs <0.5 kHz. The OC units had BFs between 2.7 and 8 kHz. To analyze the temporal responses to the complex pitch stimuli we calculated the all-order interspike-interval distribution for each unit and each stimulus condition by measuring the time between each spike and all subsequent spikes in the same spike train and tallying the intervals in an ISIH. The bin-values are expressed as firing rate (spikes per second) by dividing the raw bin count by the product of bin-width (50 μs) and the total number of spikes in response to the stimulus (Abeles, 1982; Shofner, 1999). When spikes are locked to a particular periodicity there is a peak in the ISIH at the corresponding interval. To estimate which peaks in the ISIH are related to the periodicity of the pitch-evoking signals we express the histograms as the change in firing rate relative to the response to GN presented at the same level as the pitch stimuli.
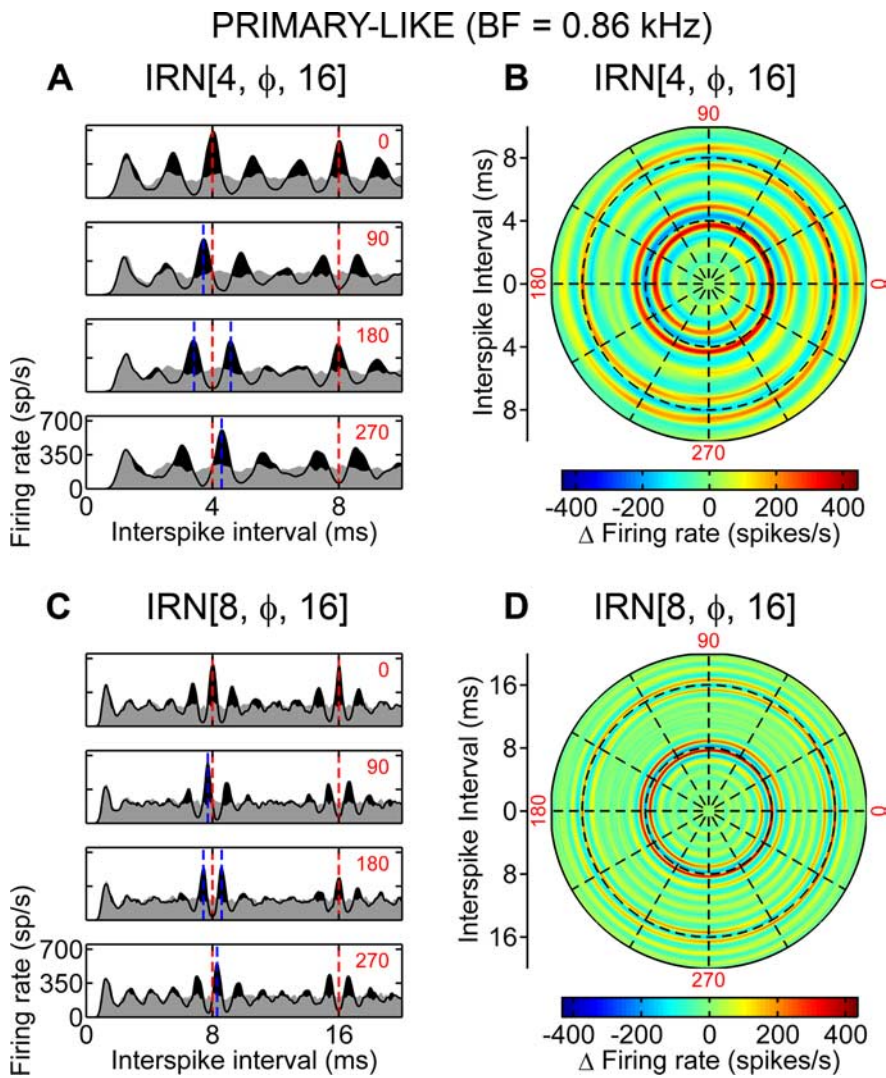
Temporal fine-structure representation

IRN is constructed by adding a broadband noise, delayed by time d and phase-shifted by $\phi°$, to the original noise and repeating this process for n iterations. The resulting signal has a series of spectral peaks with $1/d$-Hz (F0) spacing, and evokes the sensation of "repetition pitch" (Bilsen, 1966). By varying ϕ, the spectral peaks shift in frequency by $\Delta f = \phi/(360d)$, but the interpeak spacing remains $1/d$ Hz; therefore altering the waveform temporal fine structure, but not the temporal envelope. Most studies of IRN have used signals with $\phi = 0°$ and/or 180°, described by a gain parameter g, because $\phi = 180°$ is equivalent to inverting the delayed waveform ($g = -1$) and $\phi = 0°$ is equivalent to a delayed signal gain of 1. Previous studies have used the notation IRN[d,g,n], whereas here we use the alternative IRN[d, ϕ, n] (with unity gain of the delayed, phase-shifted noise). In addition to IRN stimuli, we also examine responses of VCN single units to inharmonic complex tones with their frequency components matched to the position of the spectral peaks in IRN. A frequency shift Δf is applied to each component of a harmonic complex tone; $\Delta f = \phi/(360d)$. When $\phi = 0°$, the spectral peaks are positioned at integer multiples of $1/d$ Hz and the signal has a single, unambiguous, pitch of $1/d$ Hz. In contrast, when $\phi = 180°$, the spectral peaks are at odd-integer multiples of $1/2d$ Hz, and the pitch is either well defined at $1/2d$ Hz, or it is ambiguous at $0.88/d$, $1.14/d$, and $1/2d$ Hz, depending on the signal's spectral content and, in the case of IRN, n (Bilsen and Ritsma, 1969/70, 1970; Fourcin, 1965; Bilsen, 1966; Yost et al., 1978; Raatgever and Bilsen, 1992; Yost, 1996, 1997).

In the time domain, varying ϕ changes the stimulus waveform autocorrelation function (ACF). When $\phi = 0°$, there are peaks in the waveform ACF at integer multiples of d milliseconds, with the number of ACF peaks equal to the number of iterations. In contrast, when $\phi = 180°$ there are nulls in the waveform ACF (i.e., negative correlation) at odd-integer multiples of d milliseconds, and peaks at even-integer multiples of d milliseconds. When the IRN[$d,180,n$] signal is bandpass filtered, the ACF nulls at odd-integer multiples of d milliseconds are flanked by ACF peaks, at either side of the null. The position of these flanking peaks (i.e.,

their distance from the null) is determined by the center frequency of the filter pass band. The distance from the null decreases with increasing center frequency, with the position of the peaks being given by $d \pm 1/2f_c$, where f_c is the filter center frequency. Similar relationships between d, the filter center frequency, and the position of the waveform ACF peaks exist for other values of ϕ. For example, when $\phi = 90°$, the first peak in the ACF is at $d - 1/4f_c$, and the first null is at $d + 1/4f_c$, with zero correlation at d milliseconds. In contrast to the waveform ACF, the ACF of the Hilbert envelope of IRN signals is independent of changes in ϕ (provided d remains constant), with low-amplitude peaks at integer multiples of d milliseconds. The same is true for complex tones. These features of signal processing are of importance when considering the responses of auditory neurons for two reasons. First, neurons are frequency tuned, so that they respond to a band-limited frequency range centered on the unit's BF. Second, the ability to encode the waveform fine structure decreases with increasing BF (decreasing phase-locking strength). Therefore, temporal coding of stimulus periodicity becomes gradually more dominated by the stimulus temporal-envelope periodicity (and therefore independent of ϕ) with increasing BF.

As predicted from the description of the physical acoustic signals neurons representing the fine structure of IRN[$d,0,n$] (or a harmonic complex tone with F0 = $1/d$ Hz) in their discharge patterns show peaks at integer multiples of d milliseconds in their ISIHs. In response to IRN[$d,180,n$] (or an inharmonic complex tone with all components shifted by F0/2 Hz) neurons representing the fine structure show nulls (flanked by a pair of peaks) at odd-integer multiples of d milliseconds, and peaks at even-integer multiples of d milliseconds in their ISIHs (Rhode, 1995; Cariani and Delgutte, 1996a; Shofner, 1999; Verhey and Winter, 2006; Sayles and Winter, 2007). This illustrates the equivalence between the waveform ACF and the all-order interspike-interval distribution of a neuron in the phase-locking range of BFs. The position of the ISIH peaks at either side of d milliseconds in response to IRN[$d,180,n$] depends on the interaction of unit BF and d (Sayles and Winter, 2007), and in modeling studies the position of the corresponding ACF peaks for a given value of d depends on the bandpass filter center frequency (Bilsen and Ritsma, 1969/70; Yost et al., 1978; Yost and Hill, 1979).

The interspike-interval representation of temporal fine-structure for IRN signals with $d = 4$ ms and $d = 8$ ms is illustrated for a relatively low-BF primary-like unit in Figure 2. This unit is tuned to 0.86 kHz; therefore it exhibits strong phase-locking to the fine structure at the output of the 0.86 kHz place along the basilar membrane (BM) [vector strength of ~0.8 in response to pure-tone stimulation at BF (Palmer and Russell, 1986; Winter and Palmer, 1990)]. Conventional ISIHs constructed from the responses to GN (gray shading) and IRN (black) with $\phi = 0$, 90, 180, and 270° show a series of peaks at interspike intervals related to d (but not necessarily at d ms). The position of the peaks varies with ϕ (Fig. 2A, C). When $\phi = 0°$ (top panels), the main peaks in the IRN response are at d and $2d$, with the largest peak at d milliseconds. If the largest peak in the all-order interval distribution were the pitch cue, as hypothesized by several models, this unit would indicate a pitch corresponding to the "correct" psychophysical value of $1/d$ Hz. The smaller peaks at either side of the large peaks are related to the unit's BF, so that in temporal models of pitch processing that sum temporal information (either autocorrelation magnitude or number of interspike intervals) across the BF axis, these small peaks "average out," leaving a common peak at d milliseconds in the population response, and, thus,

Figure 2. Representation of IRN temporal fine-structure by a primary-like unit. Responses from a low-BF primary-like unit (BF = 0.86 kHz). **A**, **B**, All-order interspike-interval distributions calculated from the responses to IRN[4, φ, 16]. **A**, Conventional ISIHs with responses to GN (gray) and IRN (black). Red dashed lines indicate *d* and 2*d*, and blue dashed lines indicate the positions of the response peaks predicted by equations in Bilsen and Ritsma (1969/70) (see Results, Temporal fine-structure representation). The phase shift φ applied to the delayed noise is indicated in the upper right of each panel. **B**, Continuous distribution of interspike intervals as a function of φ, plotted on polar coordinates and expressed as the change in firing rate relative to the response to GN. Radial dashed lines indicate the values of φ (30° steps) for which responses were measured. Between these values we used linear interpolation in 5° steps, before plotting the data on polar coordinates and then applying linear interpolated shading (Matlab). Circular dashed lines indicate *d* and 2*d*. **C**, **D**, As for **A** and **B** except in response to IRN[8, φ, 16]. **A–D**, Bin-width is 50 μs in all plots.

represent the unambiguous pitch of 1/*d* Hz (Meddis and Hewitt, 1991; Cariani and Delgutte, 1996b). By changing φ, and thereby altering the fine structure of the IRN signal, the positions of the ISIH peaks change so that the largest peak is no longer at *d* milliseconds. The position of the largest peak(s) in each ISIH is well predicted by Bilsen and Ritsma's (1969/70) equations relating the pitch of band-filtered rippled noise to the time interval between major peaks in the fine structure (blue lines). When φ = 180°, there is a null in the ISIH at *d* milliseconds flanked by two approximately equal-amplitude peaks, the position of which is matched by *d* ± 1/2 *BF*. On the basis of the predominant-interval hypothesis, these peaks would indicate an ambiguous pitch sensation, with one pitch just <1/*d* Hz and the other just >1/*d* Hz. Similar predictions can be made for the positions of the ISIH peaks when φ = 90 and 270°. In these conditions, the major peak
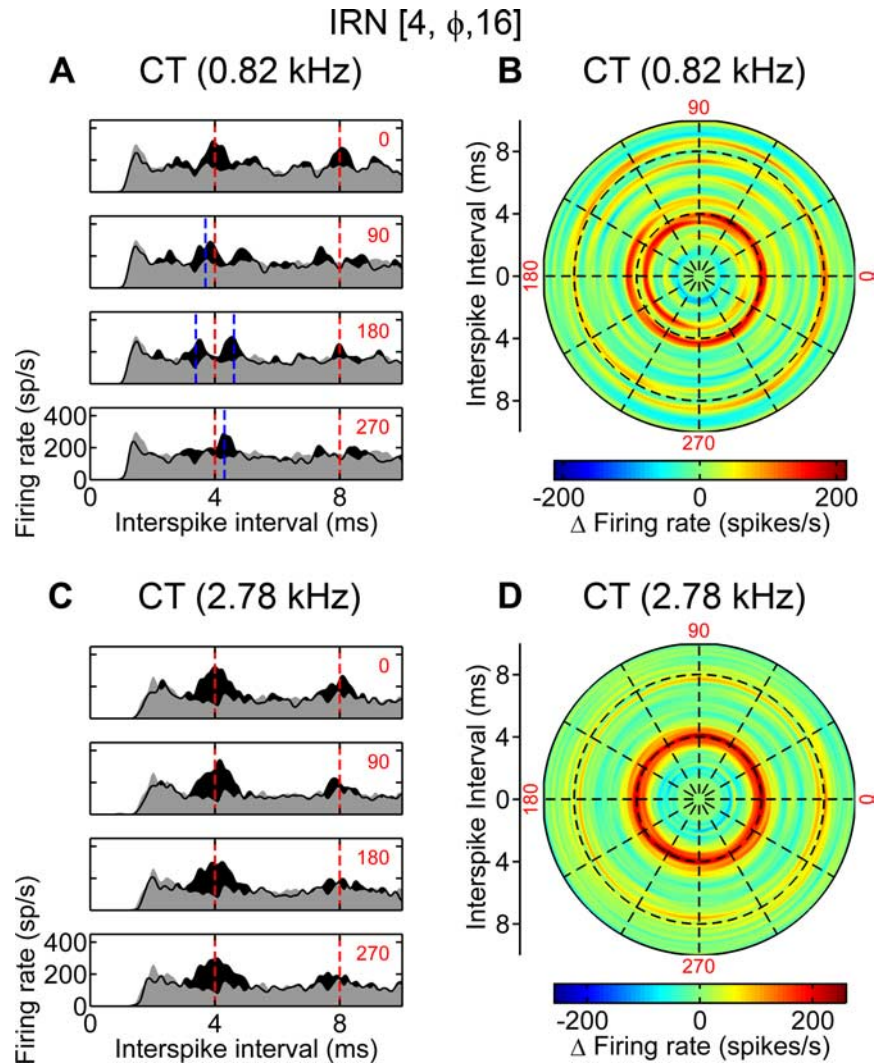
shifts away from *d* milliseconds by a small amount, indicating a small upward or downward shift in the perceived pitch, which is predicted by *d* − 1/4 *BF* and *d* + 1/4 *BF*, for φ = 90 and 270° respectively (blue lines). The circular plots (Fig. 2*B*,*D*) show the representation of stimulus fine-structure as a continuous function of φ. We measured the responses at φ = [0:30: 330], and interpolated between these points. Data are shown relative to the unit's response to GN; therefore, the color scale represents a change in firing rate as a function of interspike interval and φ. Starting at φ = 0°, as φ increases, the main peak (red) shifts toward shorter interspike intervals and the next peak at a slightly longer interspike interval becomes gradually more prominent until at φ = 180° both peaks are of approximately equal amplitude. This suggests a gradual increase in ambiguity with increasing φ up to 180°. The opposite is true when φ>180°, with the pattern of peaks suggesting a gradual shift back to the well defined pitch of 1/*d* Hz at φ = 360°.

Multipolar cells (chopper units) exhibit poorer phase-locking to pure tones than do primary-like units (Blackburn and Sachs, 1989; Winter and Palmer, 1990). Although this limits the ability of chopper units to represent stimulus fine structure in their temporal discharge pattern, if the unit BF is sufficiently low there can still be a salient fine-structure representation in these units (Louage et al., 2005; Verhey and Winter, 2006; Sayles and Winter, 2007). Responses of a low-BF chopper-T unit show a clear representation of the gradually shifting stimulus fine structure of IRN[4, φ, 16] (Fig. 3*A*,*B*). Another chopper-T unit with a BF outside the range of phase-locking for chopper units shows a relatively broad peak at *d* milliseconds independent of φ, representing the stimulus temporal-envelope modulation (Fig. 3*C*,*D*). Onset-chopper units typically

have very broad receptive fields and have previously been hypothesized as having a role in the detection of common envelope modulation across peripheral frequency channels to enhance signal detection against a modulated background (Pressnitzer et al., 2001b; Verhey et al., 2003). Responses of an onset-chopper unit (BF = 7.35 kHz) to broadband IRN, lowpass-filtered IRN and broadband COS stimuli are shown in Figure 4. If an onset-chopper unit receives only input from low-frequency channels then it is capable of following the fine-structure (Fig. 4*A*), but if it receives inputs across frequency the unit responds (weakly) to the relatively low-amplitude envelope modulation of IRN (Fig. 4*B*). The phase of modulation in IRN varies across frequency bands. Therefore it is not surprising that onset-chopper units show a relatively poor temporal representation of the modulation in

broadband IRN (Evans and Zhao, 1998). In response to COS complex tones, in which there is common modulation across frequency channels, the onset-chopper units provides a strong representation of the stimulus temporal-envelope modulation (Fig. 4C).

To determine which interspike intervals in the all-order interval ISIH were represented significantly more in response to the complex pitch-evoking signals than in the response to GN we performed a Fisher-Pitman permutation test on the interspike-interval distributions (see Materials and Methods). This non-parametric approach allows the calculation of a p value for each bin in the ISIH. We considered any bin with $p < 0.02$ to be significantly different from the response to GN. Figure 5 shows the distribution of significant peaks (localized maxima with $p < 0.02$) in the interval-difference function (the difference between the ISIHs in response to IRN and GN) as a function of harmonic number (the equivalent harmonic rank of unit BF when $\phi = 0°$) for four values of ϕ (0, 90, 180, and 270°). The data presented in Figure 5 are calculated from the responses of 45 units (34 PL/PN and 11 LF units) with BFs ≤3.5 kHz. The data are pooled from responses to IRN stimuli with d equal to 2, 4, 8, and 16 ms. Therefore, any individual unit may contribute data points to (at most) four points along the horizontal axis in each plot, with an unlimited number of points (i.e., interspike-interval distribution peaks) along the vertical axis for each point along the horizontal axis (i.e., the harmonic number corresponding to the closest integer multiple of $1/d$ to the unit's BF for any given value of d). The size of the data point is proportional to the amplitude of the peak in the interval-difference function, as indicated on the scale at the bottom of Figure 5. The dashed red lines show the predicted (normalized for d) position of the largest peak in the waveform autocorrelation function, based on the data of Bilsen and Ritsma (1969/70). When $\phi = 0°$ (Fig. 5A), the predicted peak is at a normalized interval of 1, because the largest peak in the fine structure occurs at d milliseconds. Shifting the phase of the delayed noise by 180° results in the two largest peaks in the IRN waveform fine structure being at either side of d milliseconds; their normalized position is given by $1 \pm (1/(2h))$, where h is the harmonic number at the unit (or filter) center-frequency (Fig. 5C). For ϕ of 90 and 270°, the normalized positions of the largest fine-structure peaks are given by

$1 - [1/(4h)]$ and by $1 + [1/(4h)]$, respectively. The largest peaks in the single-unit all-order interspike-interval responses follow these predictions closely. Smaller, but nevertheless significant, peaks in the ISIHs represent action potentials phase-
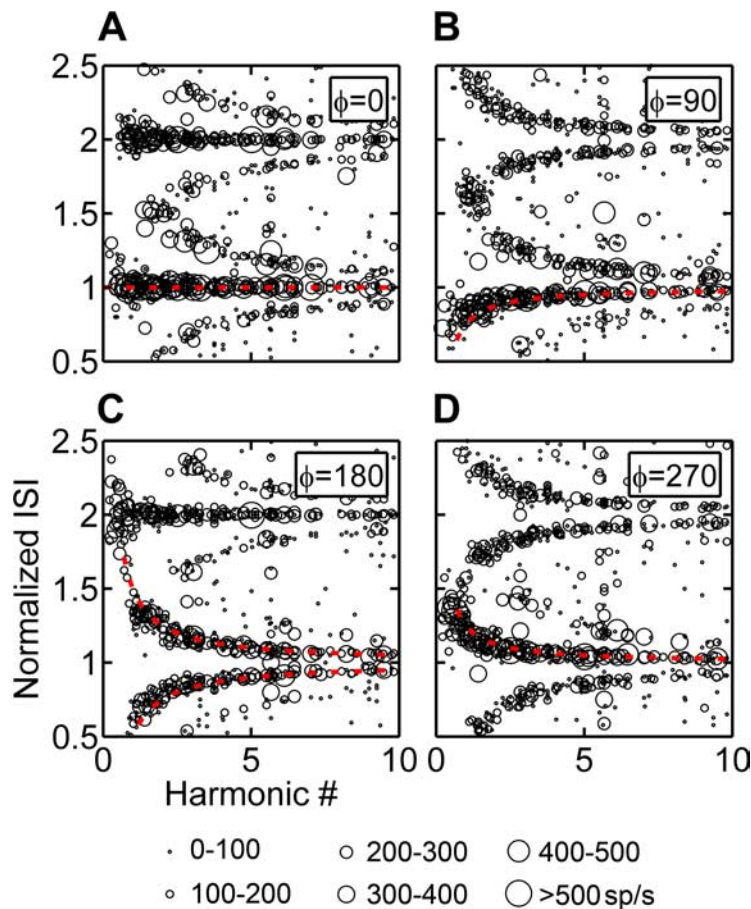
IRN [4, φ,16]

Figure 3. Representation of fine structure and temporal-envelope modulation by chopper units. **A, B,** Responses of a low-BF (0.82 kHz) chopper-T unit to IRN[4, φ, 16]. **C, D,** Responses of a relatively high-BF chopper-T unit (2.78 kHz) to IRN[4, φ, 16]. **A–D,** Binwidth is 50 μs. Circular black dashed lines are at d and $2d$ ms in each plot. The radial dashed black lines represent 30° intervals.
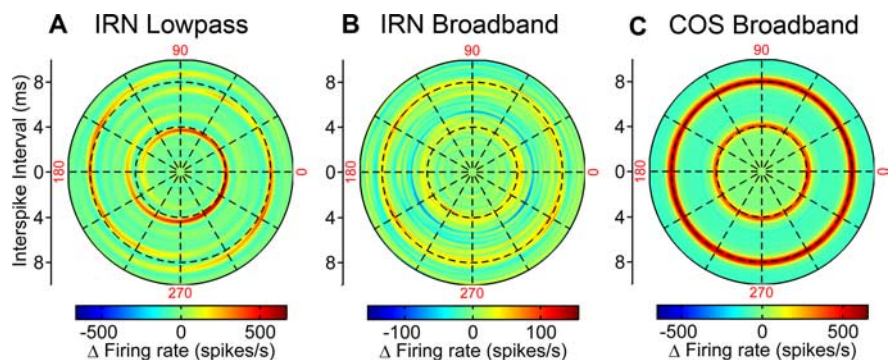
Figure 4. Responses of an onset-chopper unit to low-pass filtered IRN, broadband IRN and COS complex tones. Unit BF = 7.35 kHz. **A,** Responses to low-pass filtered (at 1 kHz) IRN[4, φ, 16] showing a representation of stimulus fine structure. **B,** Responses of the same unit to broadband (0–10 kHz) IRN[4, φ, 16] showing a weak response to temporal envelope modulation. **C,** Responses of the same unit to COS[4, φ] showing a much stronger response to envelope modulation. **A–C,** Binwidth is 50 μs. Circular black dashed lines are at d and $2d$ ms in each plot. The radial dashed black lines represent 30° intervals.

Figure 5. Significant peaks in the interspike-interval distributions accurately represent the temporal fine structure. **A–D**, Normalized interspike interval (ISI/d) at which the response to IRN is significantly elevated above the response to GN ($p < 0.02$, Fisher–Pitman permutation test) plotted as a function of harmonic number at BF. Data are from 34 PL/PN units with BFs ≤3.5 kHz, and 11 LF units. The size of the data points is proportional to the amplitude of the corresponding peak in the interval-difference function, indicated by the scale at the bottom of the figure. As described in the text, the dashed red lines indicate the predicted all-order interspike-interval distribution (temporal fine structure) peaks, based on Bilsen and Ritsma (1969/70). **A**, Responses to IRN[d,0,16] with d = 2, 4, 8, and 16 ms. Dashed red line is at a normalized interval of 1. **B**, Responses to IRN[d,90,16]. Dashed red line is at $1 - [1/(4h)]$, where h is harmonic number at BF (when $\phi = 0°$). **C**, Responses to IRN[d,180,16]. Dashed red lines are at $1 - [1/(2h)]$ and $1 + [1/(2h)]$. **D**, Responses to IRN[d,270,16]. Dashed red line is at $1 + [1/(4h)]$.

ton and Carlyon (1994), in which the number of harmonics in the 10-dB bandwidth of the filter is estimated as F0 divided by 1.8 times the equivalent rectangular bandwidth (ERB). Guinea-pig ERB is given by $0.29BF^{0.56}$ (Evans, 2001). When <2 harmonics are present in the 10-dB bandwidth of the unit's receptive field the component is said to be resolved, when >2 but <3.25 harmonics are within the receptive field the components are partially resolved, and when >3.25 harmonics are present the components are "unresolved." In the resolved condition there is little appreciable difference between the responses to COS and RAND stimuli (Fig. 6A,B). The fine-structure of both stimuli is represented in the interspike-interval distributions and there is no effect of relative component phase. The main difference between the responses to resolved stimuli is the smaller peak at intervals close to 2d ms in the IRN condition compared with the COS and RAND conditions. This reflects the noise component of IRN making the correlation in the waveform fine-structure weaker with increasing delay relative to a random phase harmonic complex (de Cheveigné, 2007). When a unit responds to several components of a complex tone (or IRN) the modulation depth and modulation rate of the temporal envelope at the output of the peripheral filter depends on the phase relationship between the components. For a COS complex tone the envelope is typically peaky, with a period corresponding to the frequency difference between the components, whereas for a RAND complex or IRN (which are physically similar) the envelope is much flatter. A peaky envelope results in the unit responding more precisely to the fine-structure periodicity in the vicinity of envelope maxima, whereas weaker envelope modulation weakens the representation of the fine structure. This influence of component phase (and therefore temporal-envelope modulation) is clear in the single-unit responses; e.g., a peak of ~750 spikes/s at d milliseconds in response to COS[8,0] is reduced to ~550 spikes/s in response to RAND[8,0] and ~450 spikes/s in response to IRN[8,0,16] (Fig. 6D–F).

Relation to the "first-effect of pitch shift"

When a harmonic complex tone is made inharmonic by the addition of a frequency shift Δf to each component, the perceived pitch shifts (Δp) in the direction of, and by an amount proportional to, Δf. This is known as "the first effect of pitch shift" and is classic evidence against the low "residue" pitch of a series of high-numbered harmonics being the result of a simple difference tone generated by cochlear distortion, because by applying an equal Δf to each component, the difference tone (and presumably any pitch perception based on the detection of it) would remain unchanged (Schouten, 1940; de Boer, 1956a,b, 1976; Schouten et al., 1962; Smoorenburg, 1970; van den Brink, 1970; Goldstein, 1973;

locked to other (smaller) peaks in the temporal fine structure (or measurement noise).

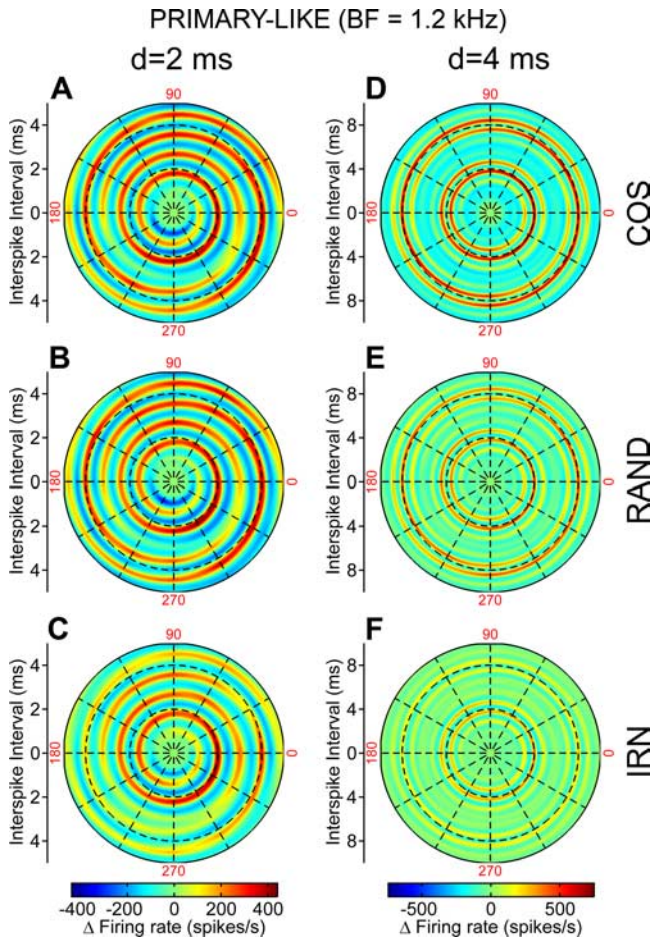**Influence of temporal-envelope modulation**

When components of a complex tone are summed in cosine (COS) phase the temporal envelope of the waveform is highly modulated with a periodicity corresponding to the intercomponent spacing, which, in the case of a harmonic complex tone, is also F0. In contrast, the temporal envelope of IRN is much less "peaky," and resembles that of a RAND phase complex tone (Winter et al., 2001; de Cheveigné, 2007). We now examine the effects of temporal-envelope modulation on the representation of fine structure by performing the same analyses on responses to COS, RAND, and IRN stimuli as a function of ϕ (see Materials and Methods). Typical results from a relatively low-BF PL unit show that the peaky envelope of the COS stimulus leads to a stronger representation of the fine-structure compared with RAND and IRN stimuli when the stimulus spectral peaks are less resolved (Fig. 6). This unit has a BF of 1.2 kHz, corresponding to the 2.4th harmonic ("resolved") when d = 2 ms (Fig. 6A–C) and the 4.8th harmonic ("partially resolved") when d = 4 ms (Fig. 6D–F). Resolvability is defined according to the rule of Shackle-

Figure 6. Effect of temporal-envelope modulation on fine-structure representation. Responses of a single primary-like unit (BF = 1.2 kHz). ***A–C***, Responses to signals with $d = 2$ ms ($\equiv$500-Hz F0). ***D–F***, Responses to signals with $d = 4$ ms ($\equiv$250-Hz F0). Top row, COS stimuli. Middle row, RAND stimuli. Bottom row, IRN stimuli. The color-scale bars beneath each column apply to all plots in that column. Binwidth is 50 μs. Circular black dashed lines are at d and $2d$ ms in each plot. The radial dashed black lines represent 30° intervals.

Figure 7. Equivalence to the "first effect" of pitch shift: single-unit responses. ***A***, Responses of a primary-like unit (BF = 1.2 kHz) to IRN with the delay set to place the third harmonic spectral peak at unit BF when $\phi = 0°$. ***B***, Same as ***A***, except with the seventh harmonic at BF. ***C***, The responses of the same primary-like unit obtained by placing the second through 10th harmonic at BF (see Materials and Methods). The data are plotted as "normalized pitch" by taking the reciprocal of the normalized ISI. The dashed black lines are the predictions of de Boer's rule for the first effect of pitch shift, $\Delta p = \Delta f/N$. Black circles are the responses of a single primary-like unit to 200% inharmonic AM tones from the study by Rhode (1995), his Figure 11*A*.

Wightman, 1973b; Patterson and Wightman, 1976; Gerson and Goldstein, 1978; Moore and Moore, 2003). Both spectral-pattern matching models and temporal models based on the detection of fine structure have been shown to account for the pitch-shift of inharmonic complex tones. By varying d so as to place the second through 10th harmonic of IRN at the BF of a unit when $\phi = 0°$ and then varying ϕ over 360° (in 30° steps), we show that the single-unit pitch matches based on the first peak in the all-order ISIH are predicted by the linear relation $\Delta p = \Delta f/N$ ("de Boer's rule" for the first effect of pitch shift), where N is the harmonic rank of the component centered on BF (Fig. **7**). We consider the effect of changing ϕ over a full cycle (360°) to be equivalent to changing the harmonic number (N) centered at the unit's BF. In the harmonic condition (when $\phi = 0°$) N is the harmonic rank of the nearest integer-multiple of $1/d$ Hz to BF (N_h). When $\phi \leq 180°$, $N = N_h + (\phi/360)$, and when $\phi \geq 180°$, $N = (N_h - 1) + (\phi/360)$.

The responses of a 1.2 kHz BF primary-like unit (not the same unit as in Fig. 6) to IRN, with the third (Fig. 7*A*) and seventh (Fig. 7*B*) harmonics positioned at BF, demonstrate the dependence of Δp on N. As the harmonic number is increased the relative deviation of the ISIH peaks from the dashed lines at d and $2d$ is decreased. By considering peaks in the all-order ISIH as "pitch

matches," we show a representation of the pitch-shift of IRN as a function of harmonic number (N) for the same single unit (Fig. 7*C*). Both single-unit responses to 200% amplitude-modulated tones (Rhode, 1995) (black circles), and predictions based on human behavioral data (Schouten et al., 1962) (black dashed lines) closely follow the single-unit responses to IRN. The general trend in the data are for a large peak in the ISIH (strong pitch representation) near to a normalized pitch of 1 at each harmonic condition, and for the peaks to become smaller (weaker pitch representation) as the complex becomes increasingly inharmonic. At many values of N, this unit predicts multiple pitch ambiguities (i.e., multiple peaks in the response for any vertical line through Fig. 7*C*). The data from Rhode (1995) in response to 200% AM tones deviate more from the linear prediction than do the data from this single unit in response to IRN. In particular, at the low harmonic numbers the pitch changes more rapidly with increasing N for the AM tone data than it does for the responses to IRN.

To compare the relative strengths of the predicted pitch matches across stimulus conditions (COS, RAND, IRN) as a function of N, we normalized the significant portions (peaks and nulls) of the interval-difference function by dividing by the maximum significant (peak) value for the COS condition for each unit and each delay condition across all values of ϕ. Thus the amplitude of the ISIH peaks are expressed relative to the response of the same unit to the COS stimulus. Typically the maximum interval-difference occurred in response to the COS stimulus at a normalized interspike-interval of 1 when $\phi = 0$ (the harmonic condition). This normalization procedure results in a normalized change in firing rate which varies between -1 and $+1$. As a function of harmonic number this analysis, for a population of 34
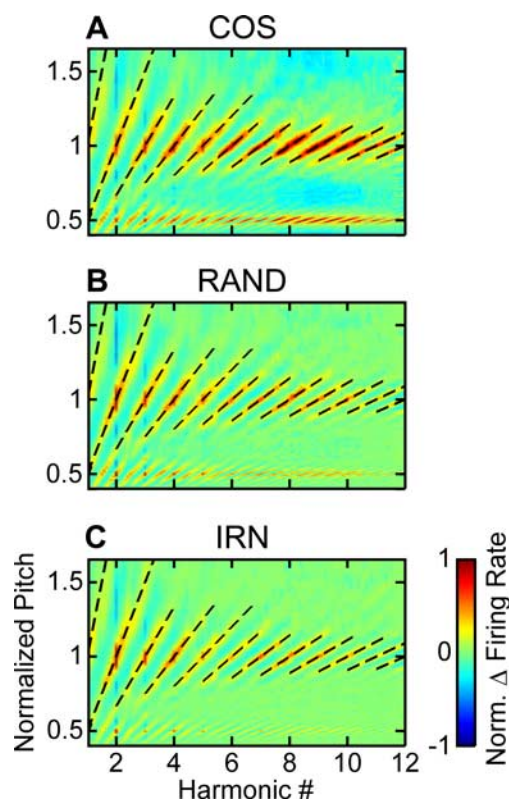
Figure 8. Equivalence to the first effect of pitch shift: Population responses to COS, RAND and IRN. *A*, "Pitch matches" based on the responses of a population of 34 PL/PN units (30 PL and 4 PN) to COS stimuli, plotted as a function of the harmonic number at BF calculated according to the rule set out in the Materials and Methods. The data plotted are the significant portions of the interval-difference function ($p < 0.02$, permutation test), normalized for each unit to the maximum peak response to COS stimulus with the same delay *d* across phase-shift conditions, and then averaged across all units. The dashed black lines are the predictions based on de Boer's rule. *B*, Same as *A*, except in response to RAND stimuli. *C*, Same as *A* and *B*, except in response to IRN stimuli.

PL/PN units with BFs ≤3.5 kHz, shows a similar pattern of results for COS, RAND, and IRN signals (Fig. 8). In each case, the pitch matches based on the all-order ISIH distribution closely follow the predictions of the first-effect (dashed lines), with the major significant peaks in the interval distribution centered at a normalized pitch of 1 when the signal is harmonic, and shifting away from 1 when the signal becomes inharmonic. When the signal is maximally inharmonic (halfway between two harmonic conditions, when $\phi = 180°$), there are two approximately equal-amplitude peaks, one indicating an upward shift in pitch and one indicating a downward shift. There are also a series of peaks in the response centered on a normalized-pitch value of 0.5. When the stimulus is maximally inharmonic these "lower-octave" peaks correspond to the true F0 of the signal, because the spectral peaks are at odd-integer multiples of $1/2d$ Hz. Lower-octave matches have been shown in human psychophysical studies using similar inharmonic signals (Gerson and Goldstein, 1978), indeed "octave errors" are often reported in behavioral pitch-matching studies with the authors choosing to "correct" these matches. Comparing the responses to COS, RAND, and IRN stimuli there is little appreciable difference between the three representations in the region of the first through fourth harmonic, because the components are resolved by the cochlear filters. In higher harmonic regions the response to COS stimuli is stronger than the response to RAND and to IRN, an effect which is consistent with

decreasing harmonic resolution and therefore greater influence of the temporal envelope of the COS stimulus on the response. For example, in the region of the eighth through 10th harmonic the response to RAND and IRN stimuli is approximately half as strong as the response to COS stimuli. Although IRN (with a large iteration number) is physically similar to a RAND complex tone (de Cheveigné, 2007) the effect of the noise component of IRN is clearly visible in the responses at interspike intervals corresponding to the lower-octave pitch matches. In this region, the response to IRN is weaker than the response to RAND complex tones, because of the cumulative effects of the (uncorrelated) noise with increasing autocorrelation delay.

The dominance region for pitch
It is well known that the lower harmonics of a complex tone are more effective at conveying pitch than are high-numbered harmonics, leading to the notion of a "spectral dominance region" for pitch (Plack and Oxenham, 2005). Although the dominance region appears to be quite variable across individuals, between experiments, and shows some dependence on F0, there is general agreement that a region around the fourth harmonic (spanning the second through fifth) dominates the pitch percept (Plomp, 1967; Ritsma, 1967; Moore et al., 1985; Dai, 2000). The existence of a dominance region for the repetition pitch of IRN has been established in human listeners (Bilsen and Ritsma, 1970; Yost et al., 1978; Leek and Summers, 2001) and in animal behavioral experiments (Shofner and Yost, 1997), and recently a physiological correlate of this has been proposed in the same species (Shofner, 2008a,b). Temporal models of pitch processing have shown that the pitch of inharmonic complex sounds can be predicted from the autocorrelation of the stimulus waveform bandpass filtered in the dominance region (Bilsen and Ritsma, 1967/68, 1969/70; Yost et al., 1978; Yost and Hill, 1979). Because the responses of auditory neurons are driven by a narrowband filtered version of the broadband stimulus waveform (by virtue of basilar-membrane filtering) the temporal responses of single neurons to inharmonic complex sounds can also be predicted from the autocorrelation of the band-filtered waveform (Sayles and Winter, 2007). Therefore there are two aspects to consider when examining the temporal representation of pitch in terms of the dominance region: the mechanism by which greater perceptual weight is applied to the dominance region, and the accuracy with which the perceived pitches are represented by neurons tuned to the dominance region. Here we present analyses examining both the preferential weighting of information in the dominance region, and the accurate temporal representation of the perceived pitch(es) by neurons in the dominance region.

Using a measure based on the relative fourth moment of IRN and GN all-order ISIHs, Shofner (2008a,b) provided evidence for the existence of a "dominance region" in the responses of Chinchilla VCN primary-like neurons to infinitely iterated rippled noise with a delay of 4 ms (IRN[4,0, ∞] in the notation used here), but did not find a similar region in the responses of chopper neurons. The aim of Shofner's study was to examine the mechanism by which greater perceptual weight is applied to the dominance region, and to establish a neurophysiological correlate of the weighting function previously identified in Chinchilla behavioral experiments. The fourth moment of a waveform is related to the variance in instantaneous power (Hartmann and Pumplin, 1988), and as used in Shofner's analysis, indicates the relative power in the ISIH measured in response to IRN, in comparison with the ISIH measured in response to GN. The ISIHs in response to IRN were renormalized, with the firing rate R_τ expressed rela-
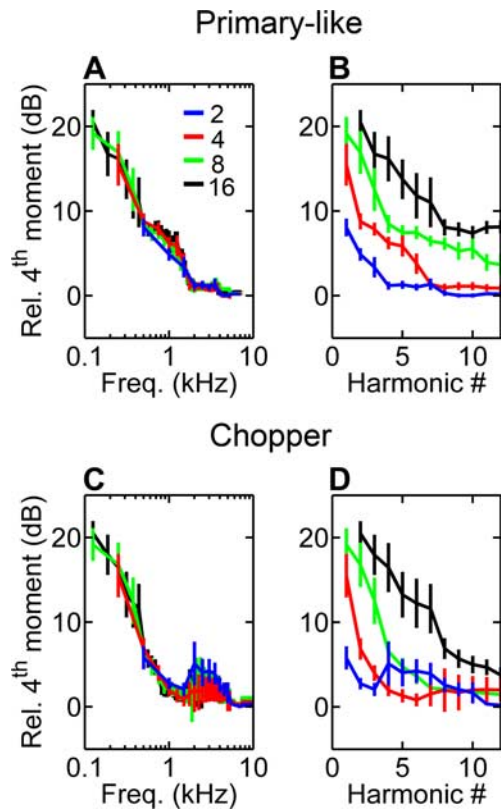
Figure 9. Spectral dominance region: relative fourth moment of the all-order ISIH. **A**, Relative fourth moment calculated for IRN[*d*,0,16], with *d* = 2, 4, 8, and 16 ms, for a population of 42 primary-like and primary-like with notch units (BF 0.23 – 6.25 kHz), and plotted as a function of unit BF with data averaged in an octave-wide sliding window with steps corresponding to 1/*d* Hz. Error bars represent ± SEM. The relative fourth moment is calculated as described in the text for a 50-ms portion of the ISIH. **B**, The same data as in **A**, plotted as a function of harmonic number. **C**, **D**, The same as **A** and **B**, but for a population of 36 chopper units (BF 0.34 –10.2 kHz). A group of 19 low-frequency units (BF 0.19 – 0.64 kHz) were allowed to contribute to both sets of data.

tive to the mean firing rate $\bar{R}$, such that the normalized firing rate $\lambda_\tau = (R_\tau - \bar{R})/\bar{R}$]. The average fourth moment, calculated over a 50 ms window, is given by $\bar{\lambda}^4 = \sum(\lambda_\tau - \bar{\lambda})^4/N$, and the relative fourth moment (dB) by dB = $10\log_{10}[\bar{\lambda}^4_{IRN}/\bar{\lambda}^4_{GN}]$.

The dominance region in Shofner's single-unit responses is correlated with the behavioral dominance region previously reported in the same species (Shofner and Yost, 1997). However, Shofner's data were limited to responses to IRN with a 4 ms delay. Therefore, it is not clear whether the physiological dominance region identified around the fourth harmonic (1 kHz) is a harmonic dominance region (i.e., the dominant harmonics would be independent of changes in *d*), or whether the region identified simply reflects the strong phase-locking of neurons tuned to the region of 1 kHz. If the latter is true, the dominant harmonics would change with *d* as the absolute frequency region of dominance would remain fixed at ~1 kHz.

We have applied Shofner's ISIH fourth-moment analysis to a population of primary-like and chopper unit responses to IRN with a range of delays (Fig. 9). In addition to the data recorded in response to IRN[*d*,0,16] (39 PL/PN, 27 CT/CS, 11 LF units) we included data recorded in response to IRN(+) from an additional 3 PL units, 9 CT units, and 8 LF units in this analysis. The analysis is shown for four different values of *d* (i.e., four different F0 s), as indicated by the color legend. In general, the relative fourth moment decreases with increasing BF for both primary-

like and chopper units in a manner consistent with the difference in phase-locking between the two unit types. It is important to realize that the relative fourth moment measures the combined response to temporal fine structure and envelope modulation; i.e., it is simply a measure of how peaky the histogram in response to IRN is, and does not distinguish between fine-structure and modulation peaks. As a function of unit BF (Fig. 9*A*,*C*), the responses of primary-like neurons decreases monotonically for all IRN delays. In the region around 0.7–1.5 kHz the primary-like units show a stronger temporal response than the chopper population, consistent with this being the region in which their phase-locking ability differs. This enhanced temporal response ~1 kHz may correspond to the dominance region identified in Shofner's analysis. Indeed, previous studies have identified neurophysiological correlates of the dominance region in the temporal responses of cat ANFs and related these to the strong phase locking ~1 kHz (Cariani and Delgutte, 1996a,b). The fall-off in phase-locking ability with increasing BF has often been associated with the fall-off in pitch salience with increasing center frequency for narrowband SAM tones, with the upper-limit of phase locking imposing an upper limit to the existence region for tonal pitch (Ritsma, 1962; Cariani and Delgutte, 1996a,b). For the chopper population (Fig. 9*C*), there is a small peak in the response between ~2 and 5 kHz, which is especially marked for IRN with *d* = 2 ms. This is likely because of a response to the envelope modulation of IRN which, at 500 Hz, is close to the intrinsic chopping frequency of many chopper units. Plotting the data as a function of harmonic number, there seems to be no evidence of a spectral dominance region tuned specifically to the fourth harmonic from a simple analysis of the relative power in the ISIH for either primary-like or chopper units.

Autocorrelation models of pitch processing are able to account for the pitch matches obtained in response to inharmonic complex sounds by assuming that the brain applies the greatest perceptual weight to the autocorrelation functions (or, equivalently, all-order interspike-interval distributions) calculated on the output of filters centered ~4/*d* Hz (Bilsen and Ritsma, 1967/68, 1969/70; Yost et al., 1978; Yost and Hill, 1979). The population analyses presented in Figure 10 examine the neural pitch-matches obtained by pooling the normalized interspike-interval distributions from all primary-like units with BFs <3.5 kHz, and from all chopper units with BFs <1.25 kHz. The unit BFs were all in the region of the second through fifth harmonics, for COS, RAND, and IRN stimuli. Data from 28 PL/PN units and from 12 CT/CS units are included in this analysis. In each plot the solid lines represent the population mean pitch-match profile for each phase-shift condition (indicated by the color code), and the shaded area around the solid lines represents the 95% confidence limits. In general, the neural pitch-matches are qualitatively similar to the human behavioral pitch-matches obtained with similar IRN stimuli [Fig. 10, compare with the data of Raatgever and Bilsen (1992), their Fig. 3]. The neural pitch matches are also similar to those predicted on the basis of the largest peaks in the autocorrelation function of the stimulus bandpass filtered in the region around the fourth harmonic.

Despite the limitations of comparing the guinea-pig neural pitch matches to human psychophysical performance, the data in Figure 10 indicate a qualitative correspondence between the overall patterns of pitch matches obtained here and in previous human psychophysical experiments such as those of Raatgever and Bilsen (1992). The primary-like unit responses to COS[*d*,0], RAND[*d*,0], and IRN[*d*,0,16] (Fig. 10*A*–*C*, blue lines) each show a large peak at a normalized-pitch value of 1 (i.e., 1/*d* Hz). This
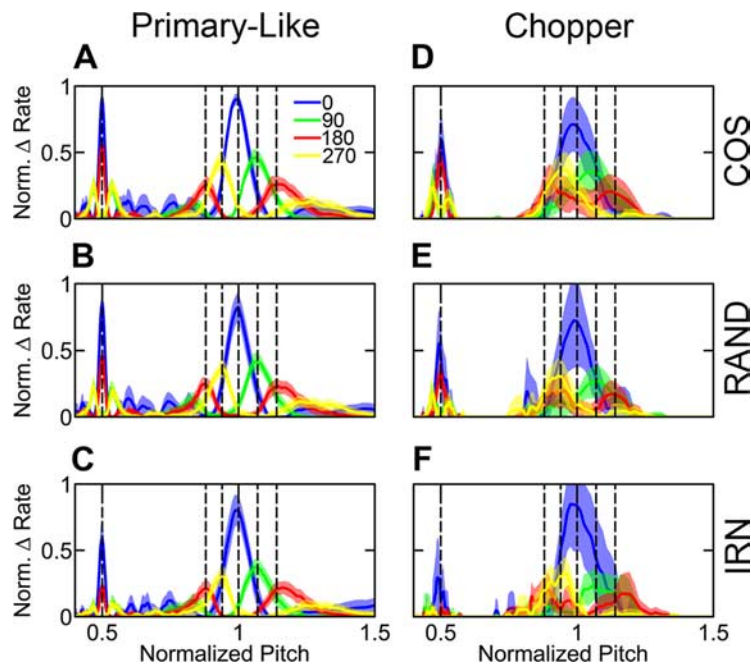
Figure 10. Neural pitch-matches in the dominance region for primary-like and chopper unit responses to COS, RAND and IRN. **A–F**, Data are the population mean pitch-match profiles calculated from the responses of 28 PL/PN units with BFs <3.5 kHz (**A–C**) and 12 CT/CS units with BFs <1.25 kHz (**D–F**). All units have BFs in the range of the second through fifth harmonic for complex sounds with $d = 2, 4, 8$, or 16 ms. **A**, The significant (positive) portions of the interval-difference function ($p < 0.02$, permutation test) in the dominance region (second through fifth harmonic) in response to COS stimuli with $\phi = 0°$ (blue), 90° (green), 180° (red), and 270° (yellow). Data are normalized to the response to the COS condition, as in Figure 8. The black dashed lines are at normalized-pitch values of 0.5, 0.88, 0.94, 1, 1.07, and 1.14, which correspond to human pitch matches for IRN stimuli (see Results, The dominance region for pitch). Solid colored lines are the mean data (across the population of units) for each condition, and the shaded area around each colored line represents the 95% confidence limits. **B**, Same as **A**, but for RAND stimuli. **C**, Same as **A** and **B**, but for IRN stimuli. **D–F**, Same as **A–C** but for a population of CT/CS units.

corresponds to the well defined unambiguous pitch of these harmonic stimuli. Shifting the spectral peaks by 1/2d Hz (red lines) the neural responses show peaks at normalized-pitch matches of ∼0.88, ∼1.14, and ∼0.5, corresponding to the perception for narrowband IRN[d,180,n] (Raatgever and Bilsen, 1992). The green and yellow lines indicate the population responses in the region of the second through fifth harmonic to stimuli with $\phi = 90°$ and 270° respectively. In both these conditions the main peak in the response is shifted away from a normalized pitch of 1, toward either a slightly higher pitch of ∼1.07 (green line, $\phi = 90°$), or a slightly lower pitch of ∼0.94 (yellow line, $\phi = 270°$). Again, these follow the predicted pitch matches based on human behavioral experiments with remarkable accuracy (Bilsen, 1966; Bilsen and Ritsma, 1969/70). The main difference between the responses to COS, RAND, and IRN is the size of the peak at ∼0.5 (i.e., the lower-octave matches). When $\phi = 180°$, this peak is largest in both the COS and RAND conditions, however, in response to IRN it is approximately equal in size to the peaks at ∼0.88 and ∼1.14. This suggests that when listening to band-filtered COS or RAND complex tones with all components shifted by F0/2 Hz, the pitch will be matched to the true F0 value more often than when listening to the equivalent IRN[d,180,n]. The neural pitch matches calculated from the responses of the chopper-unit population follow the predicted matches less accurately. By calculating the distance between the predicted pitch matches and the nearest major peaks in the neural pitch-match profiles we estimate the "error" of the neural pitch matches. The mean percentage error (±SEM) for the PL/PN population is 0.42% (±0.13%), and

for the CT/CS population 1.58% (±0.37%). For a fundamental frequency of 250 Hz ($d = 4$ ms) this corresponds to a mean time-domain error of 16 and 63 μs for PL/PN and CT/CS groups respectively.

Discussion
Temporal fine-structure
Temporal-envelope information is important for speech understanding in quiet (Shannon et al., 1995; Smith et al., 2002) and can support rudimentary pitch perception in cochlear-implant listeners (Moore and Carlyon, 2005). However, temporal-envelope periodicity is degraded by noise and by reverberation (Moore and Carlyon, 2005; Qin and Oxenham, 2005; Sayles and Winter, 2008b), and recent evidence suggests pitch perception in real environments relies heavily on fine-structure periodicity (Smith et al., 2002; Sayles and Winter, 2008b). Individuals with cochlear hearing impairment show specific deficits in their ability to make use of fine-structure cues (Lorenzi et al., 2006; Hopkins et al., 2008), and cochlear implants provide only temporal-envelope information (Wilson et al., 1991; Moore and Carlyon, 2005). Current interest in fine-structure information is therefore based on understanding the extent to which it is required for normal auditory function, the extent to which its use is limited by hearing impairment, and the development of technology capable of restoring fine-structure sensitivity.

Reports of IRN fine-structure representation in the VCN have demonstrated that primary-like units provide a more robust representation than do chopper units, consistent with the lower-limit of phase-locking in chopper units (Shofner, 1999; Verhey and Winter, 2006; Sayles and Winter, 2007). The present results show an accurate temporal representation of IRN fine-structure, with significant peaks in all-order ISIHs at intervals predicted by autocorrelation models of monaural and dichotic repetition-pitch perception acting on narrowband IRN (Bilsen and Ritsma, 1969/70; Bilsen and Goldstein, 1974; Yost et al., 1978). Thus, at the level of the VCN the fine structure representation of the pitch of inharmonic complex sounds established at the level of the auditory nerve is preserved. Because the upper-limit of phase-locking decreases as the auditory pathway is ascended, it is commonly believed that the temporal representation of pitch is transformed to a more stable rate-based representation at higher levels, probably in the inferior colliculus (Winter, 2005). It has been suggested that a key component in this transformation is the bandpass periodicity tuning of VCN chopper units in response to AM tones and to IRN (Keilson et al., 1997; Winter et al., 2001; Wiegrebe and Meddis, 2004). However, it is important to realize that this "chopper model" of pitch requires chopping periods as long as 30 ms (to account for the lower limit of pitch), whereas most VCN chopper units have an intrinsic periodicity in the range of 2–5 ms. Therefore, the contribution of VCN primary-like units in conveying the pitch-related information to higher levels should not be ignored.

Pitch-shift and pitch-ambiguity

We have shown temporal representations of the pitch-shift and pitch-ambiguity of three classes of broadband inharmonic complex sounds (COS, RAND, and IRN) rely on the use of fine-structure information in the phase-locked discharge patterns of VCN neurons. In human psychoacoustic experiments (Schouten et al., 1962) and in ANF recordings in the cat (Rhode, 1995) the pitch-shift of inharmonic SAM tones varies faster than the linear relation $\Delta p = \Delta f / N$ predicts. This deviation from linearity is known as the second effect of pitch shift, and is thought to be attributable to combination tones (Smoorenburg, 1972; Buunen et al., 1974). There appears to be no second effect in the present data. It is important to consider which factors may account for this. When AM of frequency f_m is applied to a tone, or to a narrowband noise, cochlear distortion generates a combination tone on the BM at the f_m place (Wiegrebe and Patterson, 1999). This provides a possible confound in experiments examining neural responses to AM tones, because a response at f_m could arise either by the detection of AM at the output of a high-frequency filter, or it could arise from the detection of the (relatively intense) combination tone. This is a particular criticism of studies showing neurons with BFs at or near to f_m responding to a group of high-numbered harmonics well outside of the unit's pure-tone response area (Biebel and Langner, 2002; McAlpine, 2004). The generation of combination tones in response to harmonic complex tones is dependent on simple phase relationships between components (Pressnitzer et al., 2001a; Pressnitzer and Patterson, 2001). Such phase relationships are weak or absent in IRN, and evidence suggests the distortion generated by IRN is relatively low in sound level (Yost et al., 1998). This could explain the lack of a second effect in the neural responses to IRN, but not in response to COS. Here, stimuli were presented at a relatively low sound level ($\sim$10 dB above threshold), which may account for the apparent lack of distortion. An alternative explanation is that because the data shown in Figure 8 are averaged across a population of units, any small effects of distortion may have been "averaged out," although in the single-unit data (Fig. 7) the interspike-interval distribution peaks follow the linear predictions very closely with no systematic deviation in the direction expected for the "second effect." The data from Rhode (1995) replotted in Figure 7C deviate from the linear predictions, especially at low harmonic numbers.

The temporal representation of pitch ambiguity at the level of the VCN may be viewed similarly to low-level representations of perceptual ambiguity for other stimulus parameters in audition, such as the ambiguity between percepts of one sound source and two sources when listening to pure-tone sequences (Pressnitzer and Hupé, 2006; Pressnitzer et al., 2008), and representations of perceptual ambiguity in the visual system (Tong et al., 2006). Attention and context have strong influences on both visual (Kanwisher and Wojciulik, 2000) and auditory (Fritz et al., 2007) perception. The pitch heard at any one instant when listening to inharmonic complex sounds may be the result of a process akin to the central switching mechanisms proposed for resolving perceptual ambiguity in other systems (Leopold and Logothetis, 1999) and may involve descending interaction between auditory cortex and brainstem structures such as the inferior colliculus (Winer, 2005; Nakamoto et al., 2008).

Spectral dominance

Recently, a neural correlate of the spectral dominance region for pitch has been proposed based on an analysis of the fourth moment of all-order ISIHs from the responses of VCN primary-like units to IRN (Shofner, 2008a,b). The responses of chopper units did not show a dominance region. Performing a similar analysis on the population of primary-like and chopper units here we found no evidence for a dominance region which remains fixed in terms of harmonic number for a range of fundamental frequencies. Instead the current data suggest that the dominance region identified in earlier studies may reflect a difference in phase-locking ability between primary-like and chopper units in the region of $\sim$1 kHz.

The psychophysical pitch matches obtained when using inharmonic IRN stimuli have been successfully predicted on the basis of an autocorrelation mechanism operating in the region of the fourth harmonic (Bilsen and Ritsma, 1967/68, 1969/70; Yost and Hill, 1979; Yost, 1997) and by spectral pattern-matching models operating within the same spectral region (Raatgever and Bilsen, 1992; Cohen et al., 1995). The analyses presented here show that the temporal fine-structure representation in a population of VCN units represents the correct pitch matches when restricted to a similar spectral region. Comparing the neural data (Fig. 10) to the psychophysical data presented by Raatgever and Bilsen (1992) indicates a close correspondence between the neural responses to IRN and psychophysical responses to inharmonic comb-filtered noise when both are filtered in the dominance region, with approximately equal probability of matching the pitch to $\sim$0.88/d, $\sim$1.14/d, and $\sim$1/2d Hz in both cases. The mechanism by which the "central processor" applies greater weight to the region around the fourth harmonic when computing pitch may be simply related to the fall-off in phase-locking, or involve some other more sophisticated process such as a lateral inhibitory network (Yost and Hill, 1979; Yost, 1982).

Conclusions

The temporal discharge patterns of guinea-pig VCN units provide a representation of the stimulus-waveform fine structure for inharmonic IRN and complex tones. Despite some differences between the peripheral auditory system in guinea-pigs and humans [e.g., cochlear filters may be narrower in humans (Shera et al., 2002), and the upper limit of phase-locking may differ between the two species (Palmer and Russell, 1986; Moore, 2003)], these low-level stimulus representations provide important insights into the processing of pitch-related information for inharmonic complex sounds by the mammalian auditory system. Further processing and ultimately the formation of the "pitch percept" by higher levels of the auditory system is likely to differ more across species than these peripheral representations. The fine-structure representation, based on the all-order interspike-interval distribution, predicts the pitch-shift and pitch-ambiguity of inharmonic complex sounds in line with classic theoretical and behavioral studies. Within the dominance region for pitch, the ambiguous neural pitch matches are similar to the ambiguous pitch-matches found in human behavioral experiments using similar stimuli. We conclude, tentatively, that these aspects of human (and other animals') pitch perception are mediated by similar neuro-temporal mechanisms, with an unknown contribution from higher-level processing.

References

Abeles M (1982) Quantification, smoothing, and confidence-limits for single-units histograms. J Neurosci Methods 5:317–325.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165.

Bernstein JG, Oxenham AJ (2005) An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. J Acoust Soc Am 117:3816–3831.

Berry KJ, Mielke PW Jr, Mielke HW (2002) The Fisher-Pitman permutation test: an attractive alternative to the F test. Psychol Rep 90:495–502.

Biebel UW, Langner G (2002) Evidence for interactions across frequency channels in the inferior colliculus of awake chinchilla. Hear Res 169:151–168.

Bilsen FA (1966) Repetition pitch: Monaural interaction of a sound with the repetition of the same, but phase shifted, sound. Acustica 17:295–300.

Bilsen FA, Goldstein JL (1974) Pitch of dichotically delayed noise and its possible spectral basis. J Acoust Soc Am 55:292–296.

Bilsen FA, Ritsma RJ (1967/68) Repetition pitch mediated by temporal fine structure at dominant spectral regions. Acustica 19:114–115.

Bilsen FA, Ritsma RJ (1969/70) Repetition pitch and its implication for hearing theory. Acustica 22:63–73.

Bilsen FA, Ritsma RJ (1970) Some parameters influencing the perceptibility of pitch. J Acoust Soc Am 47:469–475.

Bilsen FA, ten Kate JH, Buunen TJ, Raatgever J (1975) Responses of single units in the cochlear nucleus of the cat to cosine noise. J Acoust Soc Am 58:858–866.

Blackburn CC, Sachs MB (1989) Classification of unit types in the antero-ventral cochlear nucleus: PST histograms and regularity analysis. J Neurophysiol 62:1303–1329.

Bregman AS (1990) Auditory scene analysis. Cambridge, MA: MIT.

Buunen TJ, Festen JM, Bilsen FA, van den Brink G (1974) Phase effects in a three-component signal. J Acoust Soc Am 55:297–303.

Cariani PA, Delgutte B (1996a) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J Neurophysiol 76:1717–1734.

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

Cedolin L, Delgutte B (2005) Pitch of complex tones: rate-place and inter-spike interval representations in the auditory nerve. J Neurophysiol 94:347–362.

Cedolin L, Delgutte B (2007) Spatio-temporal representation of the pitch of complex tones in the auditory nerve. In: Hearing - from sensory processing to perception (Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S, Verhey JL, eds), pp 61–68. Berlin Heidelberg: Springer.

Cohen MA, Grossberg S, Wyse LL (1995) A spectral network model of pitch perception. J Acoust Soc Am 98:862–879.

Dai H (2000) On the relative influence of individual harmonics on pitch judgment. J Acoust Soc Am 107:953–959.

de Boer E (1956a) On the "residue" in hearing. PhD thesis, University of Amsterdam, Amsterdam, The Netherlands.

de Boer E (1956b) Pitch of inharmonic signals. Nature 178:535–536.

de Boer E (1976) On the "residue" and auditory pitch perception. In: Handbook of sensory physiology (Keidel WD, Neff WD, eds), pp 480–583. New York: Springer.

de Cheveigné A (2005) Pitch Perception Models. In: Pitch: Neural Coding and Perception (Plack CJ, Oxenham AJ, Fay RR, Popper AN, eds), pp 169–233. New York: Springer.

de Cheveigné A (2007) Comment on: "Searching for a pitch centre in human auditory cortex". In: Hearing–from sensory processing to perception (Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S, Verhey JL, eds), pp 90–91. Berlin Heidelberg: Springer.

Evans EF (2001) Latest comparisons between physiological and behavioural frequency selectivity. In: Physiological and psychophysical bases of auditory function (Breebart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonoven R, eds), pp 382–387. Maastricht: Shaker Publishing BV.

Evans EF, Zhao W (1998) Periodicity coding of the fundamental frequency of harmonic complexes: physiological and pharmacological study of onset units in the ventral cochlear nucleus. In: Psychophysical and physiological advances in hearing. Proceedings of the 11th international symposium on hearing (Palmer AR, Rees A, Summerfield AQ, Meddis R, eds). London: Whurr.

Fay RR, Yost WA, Coombs S (1983) Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system. Hear Res 12:31–55.

Fourcin AJ (1965) The pitch of noise with periodic spectral peaks. In: 5e Congres International d'Acoustique, p B42. Liége, Belgium: Société Française d'Acoustique.

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention–focusing the searchlight on sound. Curr Opin Neurobiol 17:437–455.

Gerson A, Goldstein JL (1978) Evidence for a general template in central optimal processing for pitch of complex tones. J Acoust Soc Am 63:498–510.

Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. J Acoust Soc Am 54:1496–1516.

Hartmann WM, Pumplin J (1988) Noise power fluctuations and the masking of sine signals. J Acoust Soc Am 83:2277–2289.

Hopkins K, Moore BC, Stone MA (2008) Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. J Acoust Soc Am 123:1140–1153.

Javel E (1980) Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones. J Acoust Soc Am 68:133–146.

Kanwisher N, Wojciulik E (2000) Visual attention: insights from brain imaging. Nat Rev Neurosci 1:91–100.

Keilson SE, Richards VM, Wyman BT, Young ED (1997) The representation of concurrent vowels in the cat anaesthetized ventral cochlear nucleus: Evidence for a periodicity-tagged spectral representation. J Acoust Soc Am 102:1056–1071.

Leek MR, Summers V (2001) Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners. J Acoust Soc Am 109:2944–2954.

Leopold DA, Logothetis NK (1999) Multistable phenomena: changing views in perception. Trends Cogn Sci 3:254–264.

Licklider JC (1951) A duplex theory of pitch perception. Experientia 7:128–134.

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A 103:18866–18869.

Louage DH, van der Heijden M, Joris PX (2005) Enhanced temporal response properties of anteroventral cochlear nucleus neurons to broadband noise. J Neurosci 25:1560–1570.

McAlpine D (2004) Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. J Neurophysiol 92:1295–1311.

Meddis R, Hewitt MJ (1991) Virtual pitch and phase sensitivity of a computer-model of the auditory periphery. I. Pitch identification. J Acoust Soc Am 89:2866–2882.

Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrodes. Med Biol Eng 10:662–672.

Moore BC, Carlyon RP (2005) Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Pitch: neural coding and perception (Plack CJ, Oxenham AJ, Fay RR, Popper AN, eds), pp 234–277. New York: Springer.

Moore BCJ (2003) An introduction to the psychology of hearing, 5th Ed. San Diego: Academic.

Moore BCJ, Glasberg BR, Peters RW (1985) Relative dominance of individual partials in determining the pitch of complex tones. J Acoust Soc Am 77:1853–1860.

Moore GA, Moore BCJ (2003) Perception of the low pitch of frequency-shifted complexes. J Acoust Soc Am 113:977–985.

Nakamoto KT, Jones SJ, Palmer AR (2008) Descending projections from auditory cortex modulate sensitivity in the midbrain to cues for spatial position. J Neurophysiol 99:2347–2356.

Oxenham AJ, Bernstein JG, Penagos H (2004) Correct tonotopic representation is necessary for complex pitch perception. Proc Natl Acad Sci USA 101:1421–1425.

Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells. Hear Res 24:1–15.

Palmer AR, Winter IM (1993) Coding of the fundamental-frequency of voiced speech sounds and harmonic complexes in the cochlear nerve and ventral cochlear nucleus. In: The mammalian cochlear nuclei: organization and function (Merchan MA, Juiz JM, Godfrey DA, Mugnaini E, eds), pp 373–384. New York: Plenum.

Patterson RD, Wightman FL (1976) Residue pitch as a function of component spacing. J Acoust Soc Am 59:1450–1459.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

Plack CJ, Oxenham AJ (2005) The Psychophysics of Pitch. In: Pitch: Neural

Coding and Perception (Plack CJ, Oxenham AJ, Fay RR, Popper AN, eds), pp 7–55. New York: Springer.

Plack CJ, Oxenham AJ, Fay RR, Popper AN (2005) Pitch: Neural coding and Perception: Springer New York.

Plomp R (1967) Pitch of complex tones. J Acoust Soc Am 41:1526–1533.

Pressnitzer D, Hupé JM (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr Biol 16:1351–1357.

Pressnitzer D, Patterson RD (2001) Distortion products and the perceived pitch of harmonic complex tones. In: Physiological and psychophysical bases of auditory function (Breebart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R, eds), pp 97–104. Maastricht, The Netherlands: Shaker.

Pressnitzer D, Patterson RD, Krumbholz K (2001a) The lower limit of melodic pitch. J Acoust Soc Am 109:2074–2084.

Pressnitzer D, Meddis R, Delahaye R, Winter IM (2001b) Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J Neurosci 21:6377–6386.

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128.

Qin MK, Oxenham AJ (2005) Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. Ear Hear 26:451–460.

Raatgever J, Bilsen FA (1992) The pitch of anharmonic comb filtered noise reconsidered. In: Auditory physiology and perception (Cazals Y, Horner K, Demany L, eds), pp 215–222. New York: Pergamon.

Rhode WS (1995) Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. J Acoust Soc Am 97:2414–2429.

Ritsma RJ (1962) Existence region of the tonal residue. I. J Acoust Soc Am 34:1224–1229.

Ritsma RJ (1967) Frequencies dominant in the perception of the pitch of complex sounds. J Acoust Soc Am 42:191–198.

Sayles M, Winter IM (2007) The temporal representation of the delay of dynamic iterated rippled noise with positive and negative gain by single units in the ventral cochlear nucleus. Brain Res 1171:52–66.

Sayles M, Winter IM (2008a) Neuronal representation of pitch ambiguity. J Acoust Soc Am 123:3715.

Sayles M, Winter IM (2008b) Reverberation challenges the temporal representation of the pitch of complex sounds. Neuron 58:789–801.

Schönwiesner M, Zatorre RJ (2008) Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. Exp Brain Res 187:97–105.

Schouten JF (1940) The residue and the mechanism of hearing. Proc Kon Akad Wetenschap 43:991–999.

Schouten JF, Ritsma RJ, Lopes Cardozo B (1962) Pitch of the residue. J Acoust Soc Am 34:1418–1424.

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

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Shera CA, Guinan JJ Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci U S A 99:3318–3323.

Shofner WP (1991) Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. J Acoust Soc Am 90:2450–2466.

Shofner WP (1999) Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: analysis of neural autocorrelograms. J Neurophysiol 81:2662–2674.

Shofner WP (2008a) Representation of the spectral dominance region of pitch in the steady-state, temporal discharge patterns of cochlear nucleus units. J Acoust Soc Am, in press.

Shofner WP (2008b) Temporal responses of cochlear nucleus units and the dominance region of pitch. Abstr Assoc Res Otolaryngol 31:223.

Shofner WP, Yost WA (1997) Discrimination of rippled-spectrum noise

from flat-spectrum noise by chinchillas: evidence for a spectral dominance region. Hear Res 110:15–24.

Simmons AM, Ferragamo M (1993) Periodicity extraction in the anuran auditory nerve. I. "Pitch-shift" effects. J Comp Physiol A 172:57–69.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90.

Smoorenburg GF (1970) Pitch perception of two-frequency stimuli. J Acoust Soc Am 48:924–942.

Smoorenburg GF (1972) Audibility region of combination tones. J Acoust Soc Am 52:603–614.

ten Kate JH, van Bekkum MF (1988) Synchrony-dependent autocorrelation in eighth-nerve-fiber response to rippled noise. J Acoust Soc Am 84:2092–2102.

Terhardt E (1974) Pitch, consonance, and harmony. J Acoust Soc Am 55:1061–1069.

Tong F, Meng M, Blake R (2006) Neural bases of binocular rivalry. Trends Cogn Sci 10:502–511.

van den Brink G (1970) Two experiments on pitch perception: diplacusis of harmonic AM signals and pitch of inharmonic AM signals. J Acoust Soc Am 48 [Suppl 2]:1355–1365.

Verhey JL, Winter IM (2006) The temporal representation of the delay of iterated rippled noise with positive or negative gain by chopper units in the cochlear nucleus. Hear Res 216–217:43–51.

Verhey JL, Pressnitzer D, Winter IM (2003) The psychophysics and physiology of comodulation masking release. Exp Brain Res 153:405–417.

Wiegrebe L, Meddis R (2004) The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus. J Acoust Soc Am 115:1207–1218.

Wiegrebe L, Patterson RD (1999) Quantifying the distortion products generated by amplitude-modulated noise. J Acoust Soc Am 106:2709–2718.

Wightman FL (1973a) Pattern-Transformation Model of Pitch. J Acoust Soc Am 54:407–416.

Wightman FL (1973b) Pitch and stimulus fine structure. J Acoust Soc Am 54:397–406.

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. Nature 352:236–238.

Winer JA (2005) Decoding the auditory corticofugal systems. Hear Res 207:1–9.

Winter IM (2005) The neurophysiology of pitch. In: Pitch: neural coding and perception (Plack CJ, Oxenham AJ, Fay RR, Popper AN, eds), pp 99–146. New York: Springer.

Winter IM, Palmer AR (1990) Responses of single units in the anteroventral cochlear nucleus of the guinea pig. Hear Res 44:161–178.

Winter IM, Wiegrebe L, Patterson RD (2001) The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. J Physiol 537:553–566.

Yost WA (1982) The dominance region and ripple noise pitch: a test of the peripheral weighting model. J Acoust Soc Am 72:416–425.

Yost WA (1996) Pitch of iterated rippled noise. J Acoust Soc Am 100:511–518.

Yost WA (1997) Pitch strength of iterated rippled noise when the pitch is ambiguous. J Acoust Soc Am 101:1644–1648.

Yost WA, Hill R (1979) Models of the pitch and pitch strength of ripple noise. J Acoust Soc Am 66:400–410.

Yost WA, Hill R, Perez-Falcon T (1978) Pitch and pitch discrimination of broadband signals with rippled power spectra. J Acoust Soc Am 63:1166–1175.

Yost WA, Patterson R, Sheft S (1996) A time domain description for the pitch strength of iterated rippled noise. J Acoust Soc Am 99:1066–1078.

Yost WA, Patterson R, Sheft S (1998) The role of the envelope in processing iterated rippled noise. J Acoust Soc Am 104:2349–2361.

Young ED, Robert JM, Shofner WP (1988) Regularity and latency of units in the ventral cochlear nucleus: implications for unit classification and generation of response properties. J Neurophysiol 60:1–29.