

HEARES 01535

Spectrotemporal receptive fields of neurons in cochlear nucleus of guinea pig

Ben M. Clopton and Patricia M. Backoff

Kresge Hearing Research Institute, University of Michigan, Ann Arbor, Michigan, U.S.A.

(Received 10 February 1990; accepted 3 November 1990)

Spectrotemporal receptive fields (STRFs) [Hermes et al., *Hear. Res.* 5, 147–178, 1981] for neurons in the cochlear nuclei (CN) of guinea pig were estimated. Sixteen periodic segments of bandlimited, synthesized noise evoked replicable, distinctive period histograms for spike discharges. All driven units in the major divisions of the CN having their characteristic frequency (CF) within the noise bandlimits had unique STRFs for a given intensity of noise stimulation. The STRF maximum corresponded to the unit's CF, and details of the STRF patterns differed over CN divisions and response classes derived from tonebursts. The sizes of features in STRFs from this mammal appeared significantly smaller in their temporal and spectral extents than those reported in the torus semicircularis of an amphibian and were roughly comparable to the few units reported from cat ventral CN [Eggermont et al., *Quart. Rev. Biophys.* 16, 341–414, 1983]. STRFs, as they are presently obtained, provide useful insight into some aspects of afferent processing and perhaps connectivity, but their interpretation is specific to the level of stimulation and limited by the need to choose a specific energy distribution to represent the stimulus.

STRF; Cochlear nucleus; Noise characterization

Introduction

The spectrotemporal receptive field (STRF) was proposed as a partial description of the relationship between sounds having multiple frequency components and the spike activity they evoke in an auditory neuron (Hermes et al., 1981). It is usually displayed as a surface pattern of stimulus energy over time immediately preceding a spike and across stimulus frequency. Deviations of the pattern from that expected if spike occurrence was not influenced by the stimulus, a flat surface with only stochastic variation, implicates a statistical link between a specific spectrotemporal pattern and an increased spike probability. The STRF concept evolved out of the reverse-correlation approach of de Boer (de Boer and Kuyper, 1968; de Boer and de Jongh, 1978) and was intended to supplement it by representing the second-order

stimulus influences on spike generation in auditory neurons (Eggermont et al., 1983).

The STRF has been studied primarily in neurons of the torus semicircularis of grassfrog (Epping and Eggermont, 1985a,b; van Stokkum, 1987); only a few STRFs for neurons in the CN of cat have been described (Eggermont et al., 1983). Patterns of STRFs for the grassfrog midbrain neurons extended to 50 ms before spike occurrence and covered an extensive frequency range, while the extent in cat was less than 10 ms before the spike with a frequency range of about a kHz. These observations suggest that aspects of the STRF may vary with site along the auditory pathway, with 'simpler' receptive fields existing peripherally. However, inferences about physiological and anatomical correlates are difficult to base on the available evidence.

The estimation of STRFs can be made using noise stimuli to evoke spike discharges. A spectrotemporal representation of the stimuli is then crosscorrelated with the spike train to estimate the pattern of energy fluctuations over frequency which immediately precede a spike discharge

(Eggermont et al., 1983). Interpretation is limited by the stimulus representation, usually an energy distribution, and the crosscorrelation technique which identifies linear relationships specifically. Because the approach depends on a somewhat complicated stimulus transform and because few data are available from mammals, our understanding of what the STRF conveys about the stimulus-spike relationship is incomplete. In this paper we present mammalian data from neurons in the CN of guinea pig. A more efficient method for the synthesis of noise stimuli and data collection are described. Aspects of STRF replicability and validity are considered, and problems of interpretation are discussed. Within this context we identify specific advantages and disadvantages of this characterization procedure.

Methods

Physiological preparation

Single unit responses were recorded from 26 adult pigmented guinea pigs weighing 220–400 g. They were anesthetized with an intramuscular injection of ketamine hydrochloride (Vetalar, 100 mg/kg) and xylazine (Rompun, 5 mg/kg) and supplemented regularly to maintain appropriate levels of anesthesia. The animal was placed on a DC-powered heating pad, and core temperature was maintained at 37.5 °C using feedback from a temperature probe. The skull was fixed to a rigid bar with stainless steel screws threaded into the dorsal cranium and covered with dental acrylic. The pinna and a portion of the external meatus were excised, and a Beyer DT-48 earphone in an aluminum enclosure acoustically damped with steel wool was coupled to the meatus forming a closed system. The acoustic system, sealed in the external meatus, was calibrated for its amplitude and phase response using a B and K 0.5-in microphone and a calibrated probe tube.

Glass micropipettes filled with 2 M potassium citrate were beveled to an impedance of 8–12 M Ω . The brain tissue was covered with warm 2% agar-agar in physiologic saline to prevent dessication and reduce brain pulsations. The posterior cranium was opened, and the cerebellum overlying the CN was aspirated to allow direct visual placement of the recording electrode on the surface of the DCN

or the accessible portion of the anterior VCN. The angle of entry, entry point, depth from surface obtained from the hydraulic microdrive advancing the electrode, tonotopic organization, and peristimulus time histograms (PSTHs) to tonebursts were used to estimate locus in the CN.

Electrophysiological recording

Spike waveforms were discriminated in amplitude with a level comparator to obtain maximum superimposition of triggered spike waveforms on an oscilloscope. Spike time was recorded with 10 μ s accuracy relative to the onset of each stimulus period. Histograms with bin widths of 160 μ s or greater were obtained from these records. Each histogram was based on 150 or 200 sequentially repeated stimulus periods.

Stimulus protocol

All of the stimuli were periodic, and the spike histograms corresponded to those periods except for tonebursts. The full stimulus protocol included histograms obtained from tones logarithmically swept upward or downward in frequency between 0.3 and 16 kHz (period = 163.8 ms), a rarefaction and condensation click (period = 102.4 ms), CF tone bursts at 20 to 30 dB above threshold (period = 102.4 ms), and the 16 periodic noise segments described below (period = 163.84 ms). The tonebursts were gated at the beginning of their period and had a 40 ms duration with a 5 ms rise-fall time. Gating was phase coherent over presentations in order to detect synchronous spike discharges in the histograms; the histograms were synchronized to the phase-coherent gating pulse and had a duration equal to the stimulus-gate repetition period. CFs were determined by manually adjusting the attenuation and frequency of tonebursts to estimate the minimum threshold for spike discharges monitored acoustically and visually. Noise stimuli were usually presented at 10 to 20 dB above the threshold for driving for a unit, which was evident from distinct, replicable peaks in the period histogram. Unattenuated, the noises were 71.2 dB RMS *re* 20 μ N/m²Hz^{1/2} spectrum level; most presentations were at 0 to 30 dB spectrum level. In addition, spontaneous spike rate was estimated for silence over the same duration as for noise stimulation. All of the stimuli were

synthesized and presented through a Modular Instruments M208 waveform register except for tonebursts whose gating was controlled by a 40 ms pulse from that source. Acoustic stimulation was not present during the approximately one minute interval when data were stored to disk and new waveforms were loaded into the waveform register.

Noise synthesis

The 16 independent noise stimuli used to evoke unit responses were first specified in the frequency domain. Time waveforms were then obtained using an inverse digital Fourier transform (IDFT) (Brigham, 1974). The complex Fourier coefficients for a noise spectrum $\mathbf{X}(f_k)$ (complex values are indicated with boldface) had equal, nonzero magnitudes from 299 Hz to 12207 Hz but zero magnitude outside this range. Frequencies were defined with a resolution of $f_1 = 6.1$ Hz, i.e., $f_k = kf_1$ where $0 \leq k \leq 8191$. The Nyquist frequency was thus 25 kHz at frequency sample $k = 4096$. The stimuli had nonzero spectral magnitudes for $50 \leq$

$k \leq 2000$ with zero padding for the remainder of the spectrum. The maximum frequency for a component with nonzero amplitude was thus less than half the Nyquist frequency. The phase for each $\mathbf{X}(f_k)$ was randomized (see below); values from 0 to 2π being equally likely. Each of the 16 noise stimuli had a phase spectrum which was independent of the others due to separate randomizations. Since the amplitude spectra were identical, the noises differed only in their phase spectra.

The IDFT of the discrete spectra produced analytic time waveforms, complex functions having only positive frequencies. The real parts of these waveforms were scaled to equate the RMS values between stimuli and then digitized to 12-bit amplitude resolution for presentation through the waveform register with a dwell of 20 μs for each time sample. The resulting waveforms were presented continuously resulting in periodic stimuli over 8192 samples having no transients at the points of repetition. The period $T = 1/f_1$ was 163.84 ms so that 200 periods provided 32.8 s of continuous stimulation. For later experiments an

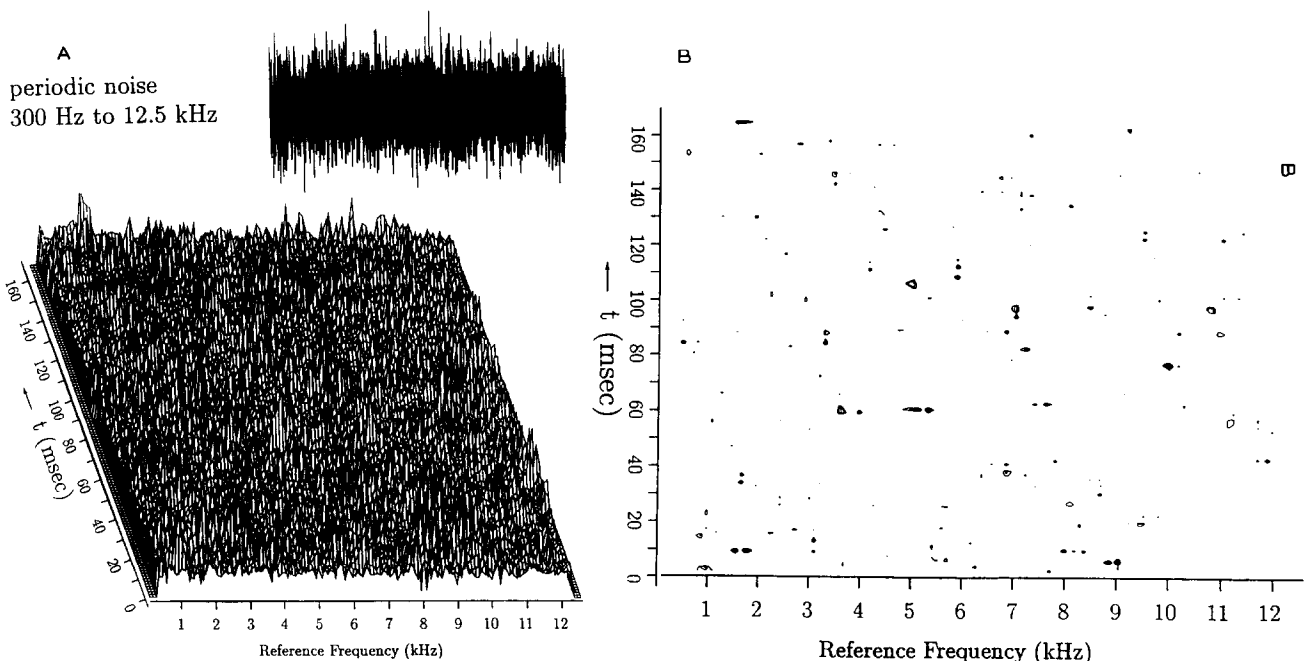


Fig. 1. Surface and contour displays of 128 by 128 values of $I(t, f_R)$ for one of the noise stimulus waveforms, shown at the top. The magnitude of $I(t, f_R)$ indicates phase coherence of the reference-frequency sinusoid, time-evolved from its Fourier coefficient, with the entire waveform at time t . Variations occur on a pedestal proportional to the power spectrum of the noise.

inverse-filter algorithm, used after the initial frequency-domain synthesis, compensated for the transfer function of the acoustic system.

Spectrotemporal stimulus representation

A transform of each noise stimulus was made to obtain a spectrotemporal stimulus matrix $I(f_R, t)$ having 128 time values (t discrete with 1.28 ms resolution) and 128 frequency values (reference frequency f_R discrete with 97.65 Hz resolution). Each noise thus had 16384 cells associated with it for this resolution, and these were stored for calculation of the STRFs. An example of one representation is shown in Fig. 1 as a surface and as a contour plot. Each cell in the matrix contained the summed values of the real part of the Rihaczek distribution (Rihaczek, 1968), equivalent to the CoSTID of Johannesma (Johannesma et al., 1981), over the cell limits. (The complex distribution is not considered here, although it was calculated). Furthermore, care was taken to prevent aliasing in the matrix. This can occur if the sample rate of the original stimulus waveform (50 kHz) is not matched to the effective sample rate of the matrix (781 Hz in this case) resulting in extraneous variance for matrix values. Details of the calculation procedure are given in the appendix.

Randomization algorithm

As described in the appendix, the detection of energy fluctuations in the noise stimuli which were related to spike discharge occurred against background fluctuations. The extent of these fluctuations was critically dependent on the starting phases of the Fourier components of the noise stimuli. Excessive phase coherence of the Fourier coefficients over a frequency region corresponds to excessive peaks of energy in the time domain. Randomization was thus critical to noise synthesis. Some of the initial results presented were obtained using a single generator algorithm to produce uniform random deviates for phase randomization. For later experiments noise stimuli were synthesized using a three-generator algorithm designed to minimize periodicities in random sequences, RAN1 from Press et al. (1986). For noise segments synthesized using both algorithms the first-order autocorrelation functions for all waveforms and crosscorrelations for waveform pairs

were effectively zero. Integration to obtain time and frequency marginal distributions of the spectrotemporal distributions indicated that the latter method reduced variance somewhat.

Data analysis

Each of the period histograms to a noise stimulus was binned at 1.28 ms and crosscorrelated successively with each of the 128 time functions in the stimulus matrix for its corresponding noise. This procedure estimated the stochastic relationship between prespike values of $I(f_R, t)$ for each f_R to give a prespike matrix or STRF. The STRF had the same cell resolution as the stimulus-transform matrix, but its time axis represented time preceding spike occurrence which we refer to as 'time precedence', τ . Thus the STRF is an estimated matrix $\hat{I}(f_R, \tau)$ which is appropriately represented as a surface. These surfaces are displayed in this paper as contour plots with the minimum contour at 2 to 3 standard deviations (σ) relative to the mean with 1σ steps beyond that unless stated otherwise. Features such as contiguous areas of high (peaks) or low (trough) values are referenced to the mean and standard deviation of the expected matrix values given no correlation between stimulus and spike probability. Major features were taken as two or more matrix values which exceeded the minimum contour criterion.

Reverse-correlation (Revcor) functions (de Boer and de Jongh, 1978) were calculated for the same noise stimuli. These functions are flat unless some synchrony exists between spike occurrence and the stimulus waveform. Histograms were binned at 20 μ s for the calculation of the Revcor functions in order to adequately sample any high-frequency synchrony that might be present.

The stability of recording was monitored visually for changes in the spike-discharge waveform, and data collection was terminated if significant changes occurred. Some period histograms were replicated and visually inspected for their similarity. A quantitative measure, the correlation coefficient for replicated period histograms, was also obtained. Correlation between binned point processes is sensitive to the bin width and spike count (Abeles, 1982; French and Holden, 1971), so all such comparisons were made with histograms having identical bins, 160 μ s, and compara-

ble total spike counts. At least two of the noise stimuli were repeated, one immediately (the 16th) and one after 10 to 15 min (the first) in order to obtain these measures.

To facilitate the identification of locus and provide a referent to previous work, unit responses to tone bursts at CF were classified into one of the major response-type categories commonly used with cat nerve and CN (Kiang, 1965; Pfeiffer, 1966; Godfrey et al., 1975a, b; Rhode and Smith, 1986a,b). The general locus of the electrode tip in

the CN could be estimated from the micromanipulator readings and the point and angle of entry.

Results

The STRFs for all units driven by the noise stimuli had discernable patterns immediately preceding spike occurrence. If no unit driving was observed, as judged from the period histograms for the noise stimuli, it was usually because the noise stimuli were presented at an intensity below

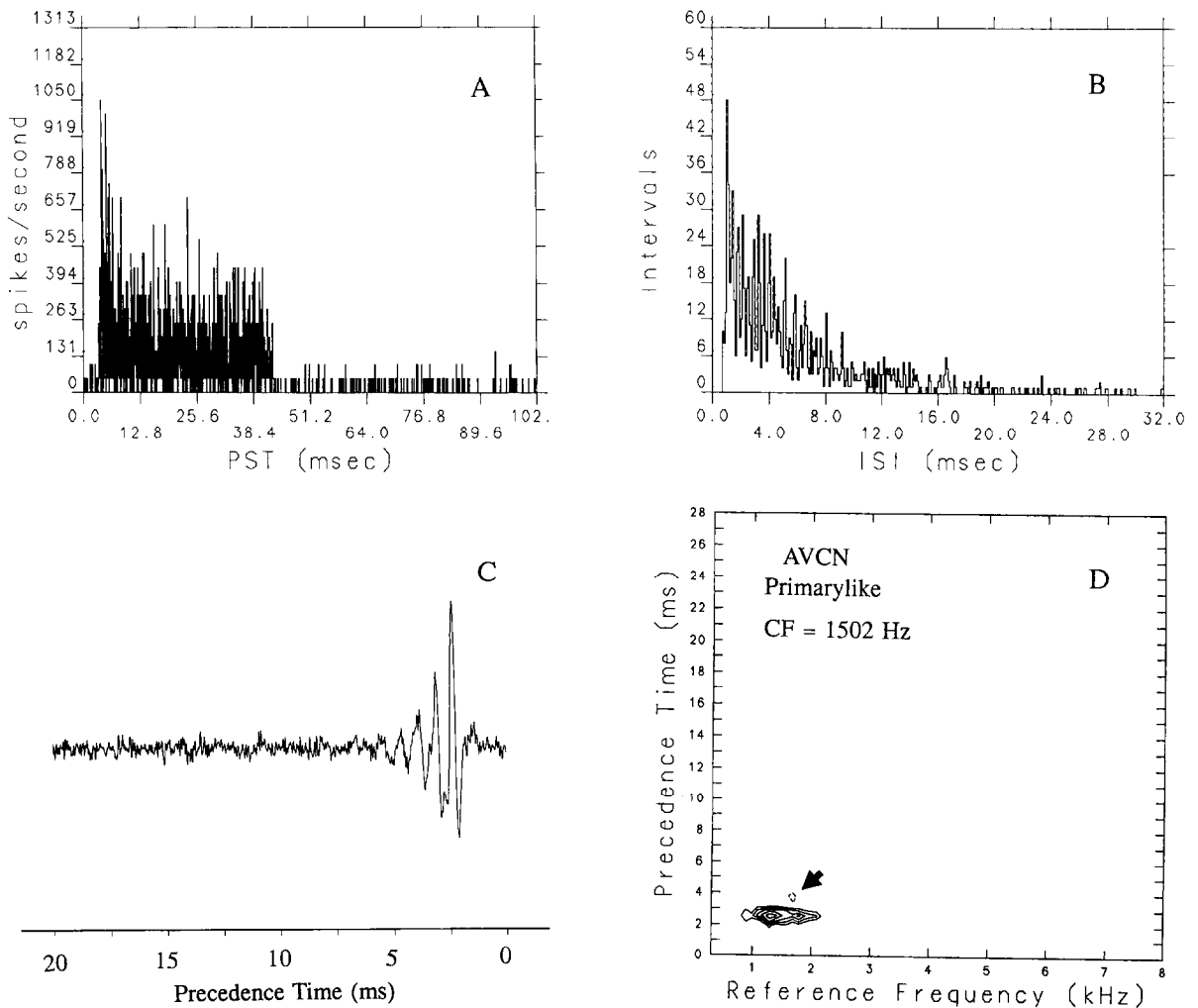


Fig. 2. STRF for a unit isolated in AVCN. (A) PSTH for 200 presentations of a 40 ms, 1502 Hz toneburst at 54 db SPL. (B) Interspike-interval histogram corresponding to (A) contains interval multiples of the CF period. (C) The Revcor function derived from the noise stimuli used to construct the STRF. (D) The STRF characterised by a major peak region and a small trough region (arrow). Based on 35,449 spikes.

threshold or the CF of the unit was in excess of 12 to 15 kHz. No pattern in the STRF was detectable in these cases. Exceptions were in the DCN, as discussed.

Examples of STRFs are shown in Figs. 2 through 9 (except for Fig. 6), the contours being chosen to illustrate STRF values which were minimally 1 to 2 standard deviations above or below the mean surface level. Maxima were generally at least 3 to 4 σ from the mean level.

Driving was obvious as a distinctive pattern of major peaks in instantaneous spike rate in period histograms derived from 200 periods of a noise stimulus. These prominent peaks replicated for a unit when the same noise was repeated, but the histogram pattern varied across units. For histograms binned at 160 μ s, the correlation between histograms which were immediately replicated was often 0.8 or higher in VCN, while replication after 10 to 15 min ranged from 0.6 upward. Units isolated in the DCN tended to have slightly lower correlation coefficients for replication, but their interpretation was complicated by lower overall spike rates for units in the DCN. While the major peaks tended to replicate well, they were sometimes time shifted by a bin between the start and end of a collection sequence. In addition, intervals of low but nonzero rates not associated with the major peaks contributed to decreased correlation between replications. This was more typical of

DCN units. Increasing the duration of the time bins for histograms often increased the quantitative measure for replication but reduced peak timing resolution through smoothing.

Results from a cell in the anteroventral CN having a primary-like response to tonebursts at CF are shown in Fig. 2. This cell demonstrated some synchrony to tonebursts at its CF, manually estimated at 1502 Hz, so it is possible to compare features of its STRF to a number of other measures. In 2A its PSTH, binned at 100 μ s, is shown to have primary-like characteristics with some indication of synchrony apparent. The synchrony is more detectable in the interspike interval histogram of 2B for CF toneburst responses. The Revcor function in 2C corroborates synchrony, but it indicates a slightly lower tuning than the measured CF since the frequency content of the Revcor function was maximal at 1380 Hz as estimated from its FFT. The Revcor function suggests that the linear portion of tuning and delay in the afferent pathway is associated with an impulse response of the shape shown. Finally, in 2D, the STRF for this cell is observed to have a single major peak region. This peak region surrounds the CF with a delay of approximately 2.58 ms, 2 time cells, which agrees with the delay of the Revcor function. The maximum in the peak region was at approximately 1370 Hz. Given a frequency resolution of 98 Hz, this approximates the CF estimate

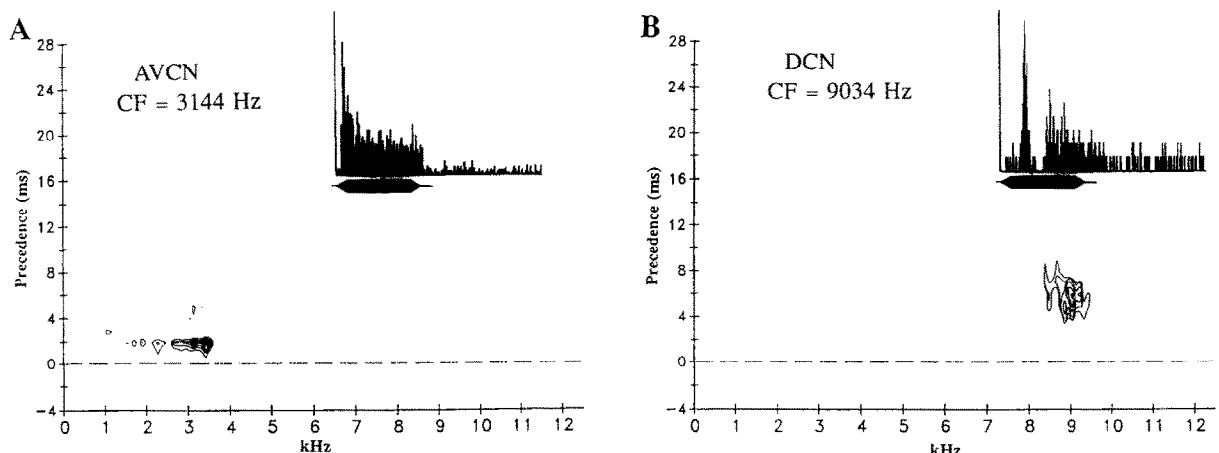


Fig. 3. Examples of two STRFs observed in the CN having primary-like (A) and pauser (B) response classifications. No evidence of synchrony to tonebursts or in the Revcor functions was present. These illustrate some of the variation between STRFs and response classifications from PSTHs for tone bursts.

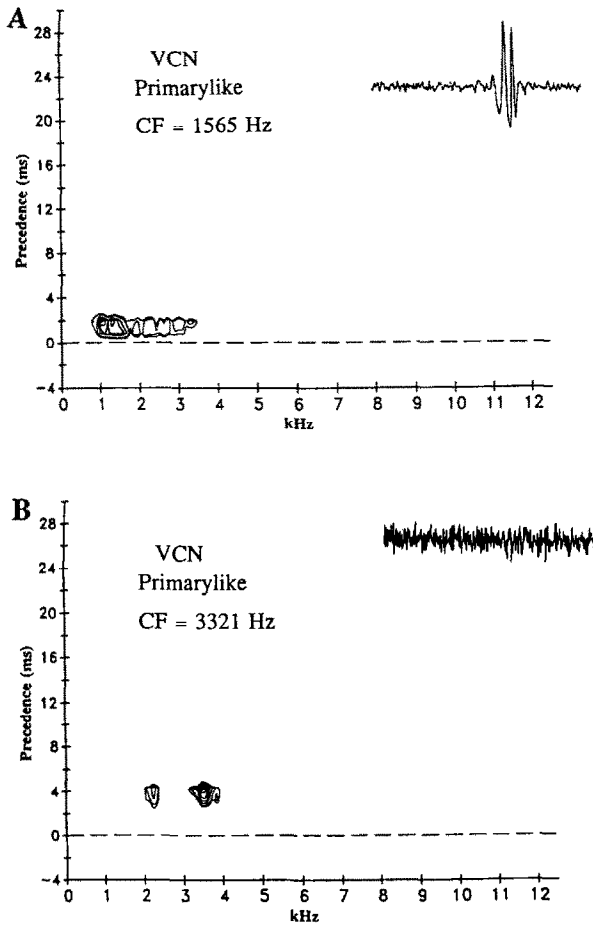


Fig. 4. STRFs for two neurons in the VCN with primary-like PSTH patterns. In (A) synchrony was present for tonebursts of lower frequencies as corroborated by the Revcor function (inset, time axis 5 ms), but no synchrony was indicated for (B). For lower contour criteria the two peaks in (B) were observed to be connected.

well and agrees with the tuning indicated by the Revcor function.

The example of Fig. 2 was typical of units with primary-like PSTHs to tonebursts except for the small trough region on the STRF indicated by the arrow. Trough regions were not commonly observed for primary-like units, although they were often seen in the STRFs of units in DCN. With some variations, primary-like units had a single band consisting of a single peak region elongated parallel to the frequency axis. The frequency range of the band invariably included the CF, and the maximum within the band was close to the CF

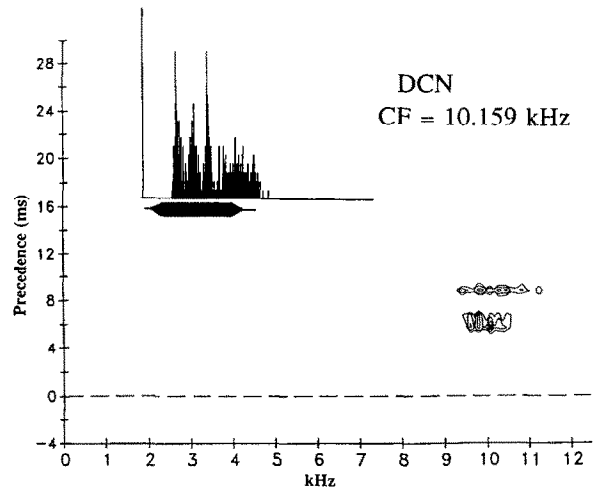


Fig. 5. STRF for a DCN unit having multiple bands and a chopper-type PSTH, shown in the inset. This double-banded peak pattern had a small trough at greater precedence (arrow).

manually estimated with tonebursts. The frequency extent of this peak region was usually 500 to 1500 Hz, and its time extent was 2–3 ms. The time precedence at the center of this major peak ranged from a few milliseconds to less than 10 ms. Given the resolution used for the calculation of these STRFs, it was difficult to more precisely estimate the dimensions of the features.

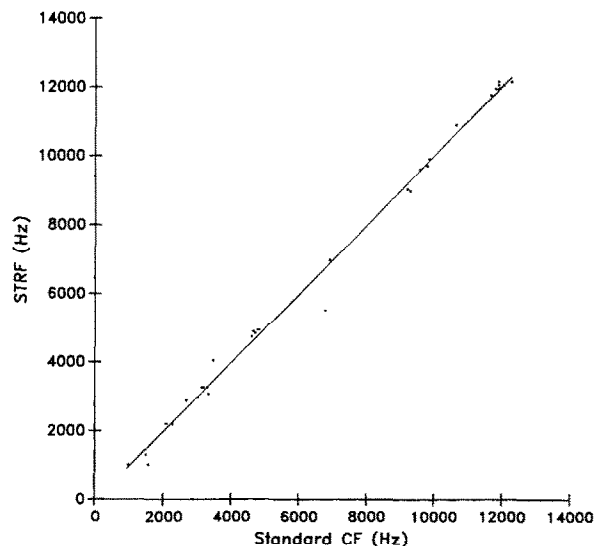


Fig. 6. Frequencies of the maximum peak on the STRF are plotted in relation to the CFs of 33 units in VCN and DCN estimated with tonebursts. The line is a least-squares regression fit with correlation coefficient 0.997.

Fig. 3 illustrates two STRF patterns for units which had no observed synchrony to tones at CF. Synchrony or its absence did not appear to influence the features of the STRF. Fig. 3A illustrates a banded peak pattern for another unit with a primary-like PSTH. In contrast, a STRF for a unit in the DCN having a pauser-type PSTH to a CF toneburst is shown in Fig. 3B. The major features differed in shape, frequency range, and time precedence, but the agreement between the maxima and the estimated CFs was retained.

Fig. 4 illustrates two units, one having significant synchrony to a toneburst at its CF and the other having none. STRFs and standard Revcor functions are shown for both. The STRFs for these two illustrate some of the variation in STRFs

for primary-like response types, but these differences were not associated with the presence or absence of synchrony. In general, we observed no difference in STRF patterns that could specifically be related to CF although functional classification with tone bursts was strongly correlated with STRF pattern. An exception to this was an increase in overall time precedence for the major peak from primary-like responders when CF was near or below 1 kHz.

Common variations on the single-peak pattern included a peak region extended in time, as shown in 3B, and a combination of peak and trough regions, as illustrated in Fig. 5. Units having chopper-type response patterns in PSTHs were most often associated with the multibanded

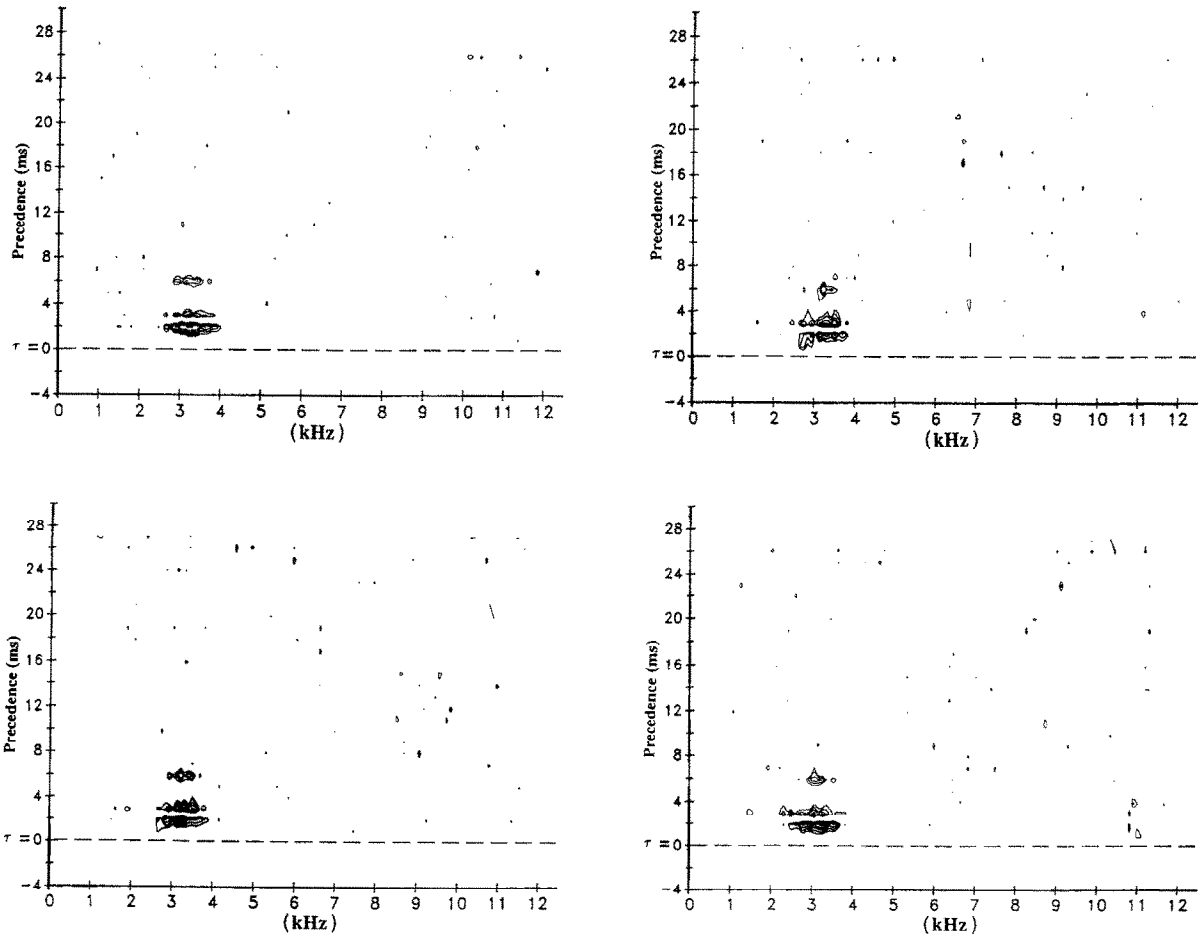


Fig. 7. Four estimates of the STRF for a DCN unit illustrating the replicability of features on STRFs. The upper pair are from the subset of noises having odd and even numbers and the lower pair are from the first and last eight of the full stimulus set.

STRFs. In Fig. 5, two linear peak regions are preceded by a small trough region. The majority of the STRFs in DCN consisted of such variant patterns. The variety of STRF patterns in the DCN are treated in detail in another paper (Backoff and Clopton, 1991).

In most cases increasing or decreasing noise intensity modified period histograms, and their corresponding STRFs underwent changes best described as variation of the major pattern. The temporal duration of histogram peaks differed between units and were as short as 100–200 μ s for some, while others were a few milliseconds. Less than 3% of the total of 84 neurons evaluated failed to give distinct STRFs to the noise stimuli. These were in DCN and were strongly inhibited by the noise stimuli. A comparison of average driven rate to the spontaneous rate was not necessarily a good indicator of driving. Decreases in driven rate for noise stimulation were observed in some instances where stable STRFs were obtained.

The STRF patterns consisted of elevations above the surface (peaks) for the sample of neurons in the VCN. Combinations of peaks and troughs (energy decrements below the surface) were observed for many neurons in the DCN (Backoff and Clopton, 1991). The peak/trough patterns invariably had positive time precedence,

that is, they closely preceded spike occurrence. Precedence values usually were in the range of 1 to 10 ms. The patterns were usually restricted in their frequency range, with a maximum near the estimated CF of the unit. The few exceptions to this involved outlying troughs which were observed for a few DCN units under conditions discussed below.

The major peak on the STRF closely approximated the CF of units, as determined manually by adjusting the frequency and intensity of repeated tonebursts, if a CF could be estimated with confidence. Some units, especially in DCN, responded to a range of frequencies making it difficult to manually 'tune' them.

The relationship between CFs estimated from tonebursts and the major peak on the STRF for 33 neurons where CF could be estimated with confidence is shown in Fig. 6. No comparison could be made if the CF fell above the upper bandlimit for the noise (12.5 kHz) since the STRF peak region could not be estimated reliably above this region. Within that range, the correspondence between the major STRF peak and the CF estimate with tones was excellent given the frequency resolution of the surfaces. STRFs for noises presented at higher levels were sometimes observed to have shifts in the frequency of the major peak.

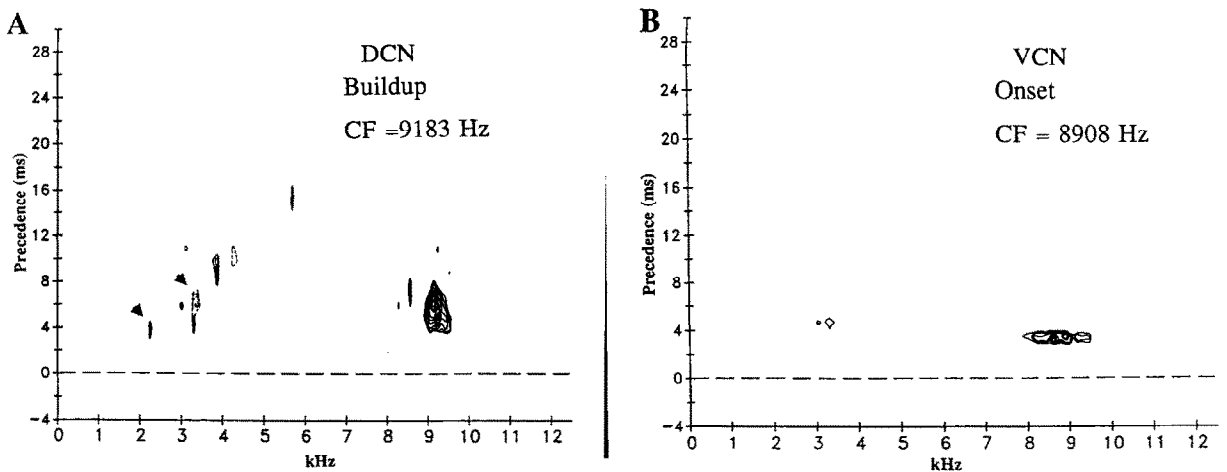


Fig. 8. Two examples of STRFs illustrating the effects of an uncompensated resonance in the acoustic system. These appear as spurious peaks or troughs in the 3 to 4 kHz region. (A) The noise series presented at approximately 25 dB above threshold for driving. While the the major peak occurred at 9231 Hz with a precedence of 6.4 ms, additional peaks and troughs (indicated by arrows) appeared in the 3–4 kHz area. (B) A major peak occurred in the 8900 Hz region near CF, but a minor peak ($+2.25\sigma$ above the mean surface) also appeared between 3–4 kHz.

The 'best frequency' for suprathreshold isointensity tones would provide a more direct comparison for tuning in those cases.

STRFs were replicated with the full noise set for a number of units with no significant change in the observed pattern. In addition, histograms for subsets of the full set provided comparable STRFs. Most units adapted to each noise in that

their average rate would tend to decrease over the 200 periods of presentation. The average rate was observed to differ over the set of 16 noises although no pattern was apparent across units. Fig. 7 illustrates the stability of a STRF pattern estimated with two independent subsets of the 16 noises. It proved possible to estimate a STRF for most units with 4–8 of the 16 noise stimuli, each

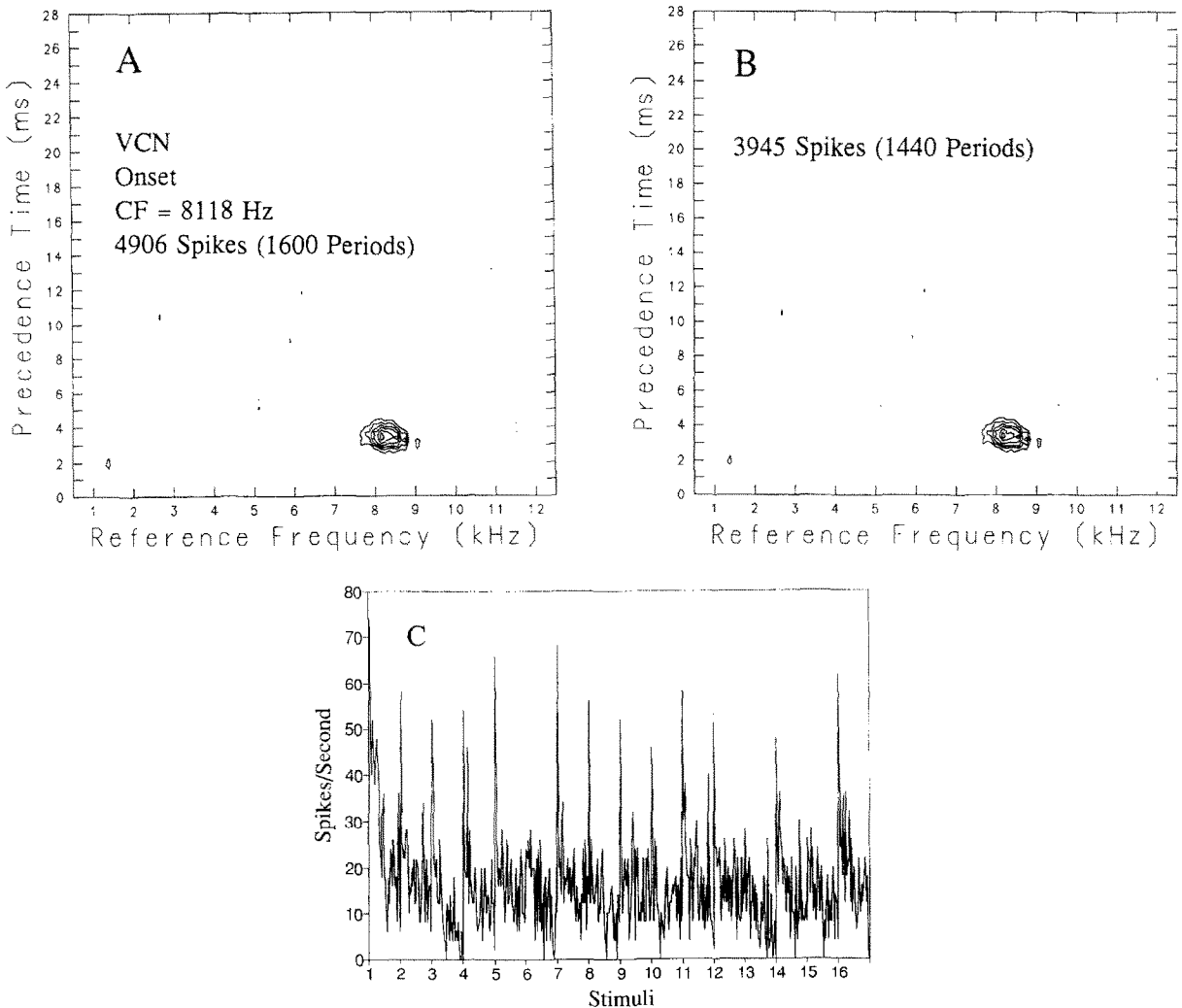


Fig. 9. Adaptation of discharge rate had little influence on the STRF illustrated in A based on 100 periods of each noise. Excluding the first 10 periods of responses from the analysis resulted in the STRF of B. A reduction of 10 percent in the number of stimulus periods resulted in a 20% reduction in the total number of spikes, but the major peak was little changed (13.6σ in B versus 13.9σ in A). The discharge rate illustrated in C was binned at 0.5 s over all of the 16 stimuli. It shows a rapid adaptation for this unit after the onset of most noises.

presented for 200 periods. Using 100 periods each of the full set also provided stable STRFs. The STRFs for units in AVCN were easily estimated with reduced stimulus sets because of their strongly patterned period histograms with high spike counts. As shown in Fig. 7 for a unit in DCN, STRFs based on a subset of stimuli tended to have more background scatter that exceeded the minimum contour levels used for revealing the major features. However, the major features replicated for all of the subsets.

Fig. 8 illustrates examples of what we interpret as artifacts observed in STRFs for units having higher CFs when noise stimuli were presented at intensities in excess of 30 dB SPL spectrum level in a system without compensation for noises generated with the initial randomization procedure. Spurious peak and trough regions were sometimes observed due, we assume, to resonances of the acoustic system in the 3 to 4 kHz range which may have driven units with high CFs through the 'tails' of their tuning curves. In these early samples, these feature separations were not affected by changes in contour criteria as in 4B.

Varying degrees of adaptation in discharge rate were observed during the 16 noise sequences. Some units had no noticeable adaptation, but most had a decrement in discharge rate that was usually detectable during data collection. The most common form of this adaptation was an exponentially asymptotic approach to a steady state over the 100 or 200 periods of continual noise presentation. In order to determine if different response properties were associated with intervals of rapid adaptation and later, more stable periods of discharge, STRFs excluding discharges during an initial period of rapid adaptation were compared to STRFs based on all of the discharges. Fig. 9 illustrates one case. The STRF based on the full data record is shown in 9A while 9B illustrates the STRF obtained when the data from the first 10 periods (1.68 s) of 100 periods were not used in the analysis. The STRFs are almost identical. The full time record for all of the noise presentations is shown in 9C to illustrate the relatively rapid drop in discharge rate at the onset of the 16.8 s for most noise presentations. During stimulation this unit had an average discharge rate of 18.8 spikes/s. Panel 9C does not include the intervals of approximately 1

min between noise sequences when no stimulation was present and recovery from adaptation occurred.

Discussion

These results indicate that distinctive, replicable patterns of energy, as measured by the real part of the Rihaczek distribution are stochastically associated with spike occurrence in the CN of guinea pig. While STRFs have been reported for a few units in cat CN (Eggermont et al., 1983), these are the first STRF estimates for units over the major CN subdivisions in mammal.

Although previous investigations have included estimates of STRFs using noise stimuli, the methods used in this study offer specific advantages. These include greater control over the statistical properties of the noise stimuli, more reliable estimates of spike-discharge probabilities, and the possibility of more extensive analyses at a later time. The noise stimuli were synthesised using the inverse discrete Fourier transform resulting in periodic stimuli which were acoustically 'seamless.' Since randomization of the phase spectrum in the frequency domain is the critical operation determining the noise properties, the method we used is dependent only on the algorithm used to generate uniform random deviates. Algorithms for this purpose are well understood (see Press et al., 1986) and less likely than hardware-generated noise to contain subtle patterning in the time or frequency domains (Swerup, 1978; Eckhorn and Popel, 1979) which could bias STRF estimates. In addition, the periodic stimuli can be presented repeatedly in order to obtain period histograms of spike discharges which better estimate discharge probability on which the STRF estimate is based. Thus, instead of presenting a nonrepeating noise sequence of long duration, statistical accuracy is increased by stimulus repetition. Finally, since the resulting ensemble of stimuli and period histograms remain available, various comparisons (an example is illustrated in Figs. 7 and 9) can be made in order to verify statistical properties of the STRF estimates, and other stimulus representations can be used to estimate STRFs (Eggermont and Smith, 1990).

The STRF patterns observed were restricted in

frequency and time relative to spike occurrence, and they were replicable and specific to a unit. The maximum peak for a unit's STRF at lower intensities agreed well with its CF determined from tonebursts, and this correspondence was accurate over the bandwidth of the noise stimuli. STRF estimates did not depend on synchrony or on an increase in the average rate of firing. As discussed in previous work and in the appendix, STRF peaks signal a brief elevation in the level of phase coherence between pairs of frequency components, one being the reference frequency, immediately preceding a spike discharge. Since tonebursts are, in effect, isolated temporal segments of phase coherence, it is reasonable to expect that the CF would be associated with the frequency at which the neuron is most sensitive to phase coherence.

Peak regions in the STRF for tonebursts are likely to reflect excitatory input over direct afferent pathways, so the time precedence for the major peak region is an obvious correlate of response latency. System delay has been shown to be reduced for continuous noise stimulation as compared to tonebursts (Møller, 1981), so the minimum time precedence is likely to approximate the minimum afferent delay. It differed for unit responses from different divisions of the CN and across frequency in qualitative agreement with expected delays from cochlear mechanics and intranuclear transmission (Backoff and Clopton, 1991).

It is less obvious what trough regions indicate. They may arise from more indirect, inhibitory pathways. They almost invariably had greater time precedence than an initial peak, and were more often observed in the STRFs of units in DCN (Backoff and Clopton, 1991). However, an hypothesis of inhibitory pathways is not supported by more extensive trough features as might be expected from toneburst data. One alternative explanation is that peak-trough patterns in STRFs arise from an enhanced spike probability following energy fluctuations which match a neuron's intrinsic oscillatory tendencies (Kim et al., 1990).

The lack of troughs lateral in frequency to major peaks, insofar as they may be correlates of lateral suppression or inhibition, is unexpected from the extensive body of work in the nerve and

CN using tones and multitone complexes (Sachs and Kiang, 1968; Deng et al., 1987; Young and Brownell, 1976). A reduction in inhibition due to anesthetic effects may have contributed to these effects (Evans and Nelson, 1973; Young and Brownell, 1976). However, an important characteristic of the noise stimuli used in this study was a flat amplitude spectrum, i.e., a lack of spectral contours within its bandlimits. Spectral contours have been characteristic of stimuli demonstrating suppression and inhibitory areas in response maps. Taken together, the evidence presented here suggests that responses in CN to wideband noise with a flat amplitude spectrum lack major inhibitory influences. This property may serve as a useful reference against which the effects of added spectral components can be evaluated. For example, changes induced in STRFs due to the presence of continuous tones of selected frequencies and intensities should provide further insight into mechanisms of suppression or inhibition. Preliminary observations indicate that continuous tones near CF added to the noise stimuli decrease and otherwise modify STRF peaks in our preparation (Clopton et al., 1990) supporting the spectral contour explanation.

The resolutions in time and frequency with which STRFs are calculated depend on theoretical and practical considerations. Eggermont and Smith (1990) have pointed out that the resolutions for integrating energy from an afferent filtering process with effective bandwidth Δf and time (envelope-fluctuation) resolution Δt requires that $\Delta t \Delta f \geq 1/4\pi$. This limit arises because the frequency and time extents of an energy fluctuation in the time or frequency domains are related by this relationship, however Cohen (1989) argues that the relationship does not limit details in time-frequency distributions. While our stimulus representations (96 Hz by 1.28 ms) meet the above criterion and were adequate to identify major STRF features, it is apparent that some were only minimally resolved. Higher resolutions are necessary to address many questions of physiological importance. The computational limit for resolution is determined by the original stimulus waveform. For the noise stimuli used here, it is 20 μ s by 6.1 Hz, but to achieve this resolution computation time increases significantly (see Appendix).

Resolution in the time domain is the major determinant of this. Increased resolution is needed not only to obtain more accurate estimates of a feature's time and frequency extent but also to reduce the possibility of cancellation of terms which are summed within a cell of the STRF. To the extent that a mixture of terms representing the phase-coherent condition are summed with terms carrying unrelated phase information, or even terms of lower than expected energy due to incoherent phases, the stimulus surface and STRFs derived from it will contain obscuring variance. This would reduce the apparent correlation between the stimulus and spike generation. Insufficient resolution is thus an obvious source of variance which might contribute to degradation of the estimate for the STRF surface.

The real parts of the spectrotemporal terms of the Rihaczek distribution indicate the degree to which the reference-frequency component is in phase with other components of the stimulus at a point in time. As outlined in the Appendix, if a linear bandpass filter, approximating cochlear filtering, receives an input waveform in which the frequency components are simultaneously phase coherent over its bandlimits, its response will tend to be of greater than average magnitude after a delay characteristic of the filter. However, the maximum response from the filter (maximum coherence at its output) is obtained when the pattern of input phase coherences match the filter's phase characteristics. The STRF, as presently obtained, reflects occurrences of brief periods of simultaneous phase coherence in the stimulus for a restricted range of frequency components preceding spikes, and this is not necessarily the optimal pattern for the cochlear filtering. This suggests that the Rihaczek is not optimal as a stimulus representation on which to base STRF calculations unless a relatively constant group delay function holds for the afferent pathway. Observations in the VIIIth nerve indicate that, at higher frequencies, at least, cochlear filtering is characterized by a constant group delay (Allen, 1983).

We stress the distinction between the STRF as a conceptual approach to encoding in the auditory system and specific implementations. The implementation must make assumptions about how to best represent complex stimuli and to estimate

important parameters of the afferent auditory system. As pointed out, the Rihaczek distribution provides a specific summary statistic of phase coherences in the stimulus. However, it is only one of a class of energy-density distributions (see Cohen, 1989 for a review). A choice better suited to the estimation of system parameters may exist within the class (e.g., Hermes, 1985; Choi and Williams, 1989; Eggermont and Smith, 1990). All of these distributions address second-order stimulus properties, so strategies aimed at the more general problem of identifying nonlinear systems (e.g., Bendat, 1990) must be used if a comprehensive system characterization is the goal. The utility of any approach is predicated on understanding its results in terms of established properties of the auditory system. Finally, any approach using wideband, flat-amplitude noise must be qualified by the possibility that it represents a unique input to peripheral auditory processing mechanisms.

In summary, the STRFs of neurons in the mammalian CN indicate tuning and delay in the afferent pathway. As presently implemented, they do not appear to indicate lateral suppression or inhibition which have been observed with multiple-tone stimulation. However, the STRF patterns are stable and replicable, and they have some correlates with functional classifications based on responses to tonebursts. The synthesis technique used in this study defined the noise stimuli in the frequency domain opening the possibility for precise experimental manipulation and generalization to a wider class of important sounds.

Appendix

Stimulus Synthesis

Each noise segment consisted of $N = 8192$ discrete samples at $20 \mu\text{s}$ intervals and was obtained from its complex spectrum using an inverse FFT. The amplitudes of the Fourier coefficients $X(f_k)$ were equal within the bandlimits of 0.299 to 12.207 kHz (considered unity for convenience). This corresponded to 1951 frequency components spaced at 6.1 Hz ($f_k = k \cdot 6.1 \text{ Hz}$ for $k = 50$ to 2000). Their phases ϕ_k were randomized so that a flat probability distribution was approximated for $0 \leq \phi < 2\pi$.

The stimulus waveform was proportional to the real part of

$$\mathbf{x}(t) = 1/N \sum_k e^{i[2\pi f_k t + \phi_k]}$$

where $i = \sqrt{-1}$.

Second-order representation of stimulus

The distribution $I(t, f_R)$ is the spectrotemporal stimulus representation used for the noise stimuli described. In its complex form it corresponds to Rihaczek's 'complex energy distribution function' (1968) and the 'coherent spectrotemporal intensity density' (CoSTID) of Johannesma et al. (1981). It is one of a class of distributions described by Cohen (1966, 1989) which partition the energy in a waveform in time and frequency.

Terms contributing to this representation arise from the complex sinusoidal components for a time-evolved reference frequency f_R and for other frequencies f_k . They are scaled by $1/N$ and time-translated to $t = j\Delta t$ where j is an integer ($0 \leq j < N - 1$) and $\Delta t = 20 \mu\text{s}$ for our noise stimuli:

$$\mathbf{I}(t, f_R, f_k) = 1/N^2 [\mathbf{X}(f_R) e^{i2\pi f_R t}]^* \mathbf{X}(f_k) e^{i2\pi f_k t}$$

where * indicates the complex conjugate. The real part of this complex product is

$$I(t, f_R, f_k) = 1/N^2 \cos[2\pi(f_k - f_R)t + (\phi_k - \phi_R)]. \quad (1)$$

It is positive if the f_R and f_k terms are in phase or 'phase coherent' at time t and negative if they are out of phase. The imaginary part of the product is necessary to specify their relative phase but is not used here. Properties necessary for a valid energy representation are retained (Cohen, 1989). An additive combination of these terms over all frequencies represents the phase coherence for the reference-frequency term with the frequency components of the stimulus waveform at time t ,

$$I(t, f_R) = \sum_{f_k} I(t, f_R, f_k),$$

and this provides a basis for generating the spectrotemporal stimulus representation used. Energy

in the time waveform at t is given by the sum over all reference frequencies,

$$I(t) = \sum_{f_R} I(t, f_R).$$

The positive square root of $I(t)$ is the envelope of $x(t)$, and the power spectrum is given by

$$|\mathbf{X}(f_R)|^2 = \sum_t I(t, f_R).$$

A complete representation of $I(t, f_R)$ would require a surface having a resolution of $20 \mu\text{s}$ by 6.1 Hz per matrix cell, at least 16 million complex values (2048×8192). Since this was beyond our computational and storage capacities, the surface was calculated with a resolution of 1.28 ms by 97.67 Hz so that the surface contained 16384 elements (128×128). A time waveform and its surface and contour displays for one of the noise stimuli are shown in Fig. 1.

Precautions against aliasing

Since the real time waveform $I(t, f_R)$ is the sum of sinusoids with frequencies ranging from zero ($f_R = f_k$) to values limited by the bandwidth of the noise, precautions must be observed to insure that the effective sampling rate for the stimulus-representation matrix is adequate to prevent aliasing (Brigham, 1974). Since the spectral coefficients for each of these waveforms were calculated and the waveform obtained with the inverse FFT, it was possible to exclude difference frequencies which were too great to be represented. In the present case, the representation was sampled at 781 Hz or $1/64$ th the rate of the original waveform, so difference frequencies in excess of 390 Hz were excluded from the spectra of each $I(t, f_R)$ to prevent aliasing.

Estimation of afferent filtering

If the input and output waveforms of a linear filtering process are represented as described above, their corresponding terms for Equation 1 differ in a manner dependent on the transfer function of the filter. If the steady-state response

of the filter is $a_k \cos[2\pi f_k t + \theta_k]$ for a cosine input, then the output terms of Equation 1 are

$$a_k a_R \cos[2\pi(f_k - f_R)t + (\phi_k + \theta_k) - (\phi_R + \theta_R)]. \quad (2)$$

Assuming that spike-probability maxima arise due to maxima in the sum of these terms, i.e., peaks in energy at the filter *output*, then spike occurrence will be associated with the cosine conditions

$$\phi_k + \theta_k - \phi_R - \theta_R = 0.$$

The input conditions which match filter characteristics producing this phase cancellation occur when $\tau_g = -\tau_{in}$ where

$$\tau_g = (\theta_k - \theta_R)/2\pi(f_k - f_R)$$

for the filter and

$$\tau_{in} = (\phi_k - \phi_R)/2\pi(f_k - f_R).$$

for the input. $-\tau_g$ approximates the filter's group delay, the negative derivative of the phase function, at f_k ($f_k > f_R$) for small separations of f_k and f_R . These considerations imply that maxima in the energy output of the filter, and presumably spike-discharge rate, will occur at a delay of τ_g after a number of Fourier components are phase coherent over a frequency region where the filtering process has a relatively flat group delay function. Thus STRFs based on the Rihaczek distribution emphasize specific phase conditions in stimulus waveforms.

In addition, the terms of Equation 1 do not differentiate between $f_k > f_R$ and $f_k < f_R$. Symmetrical summation around f_R will obscure the unique contribution of specific pairs of frequencies. This is a property common to Cohen's class of distributions.

Acknowledgements

This work was supported by NIH grant NS05785 and equipment grant BNS8609850 from the NSF. We thank Daryl Kipke for help with data analysis.

References

- Abeles, M. (1982) Quantification, smoothing, and confidence limits for single-units' histograms. *J. Neurosci. Meth.* 5, 317-325.
- Allen, J.B. (1983) Magnitude and phase-frequency response to single tones in the auditory nerve. *J. Acoust. Soc. Am.* 73, 2071-2092.
- Backoff, P.M. and Clopton, B.M. (1991) A spectrotemporal analysis of DCN single unit responses to wideband noise. *Hear. Res.* 53 (in press).
- Bendat, J.S. (1990) *Nonlinear System Analysis and Identification from Random Data*. Wiley, New York.
- Brigham, E.O. (1974) *The Fast Fourier Transform*. Prentice-Hall, Englewood Cliffs, N.J.
- Choi, H.-I. and Williams, W.J. (1989) Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Trans. Acoust. Speech, Signal Processing* 37, 862-871.
- Clopton, B.M., Backoff, P.M. and Kipke, D.R. (1990) System estimation and response prediction from high-resolution spectrotemporal receptive fields in cochlear nucleus. *Abstr. Assoc. Res. Otolaryngol.* 337.
- Cohen, L. (1966) Generalized phase-space distribution functions. *Math. Phys.* 7, 781-786.
- Cohen, L. (1989) Time-frequency distributions—A review. *IEEE Proc.* 77, 941-981.
- de Boer, E. and Kuyper, P. (1968) Triggered correlation. *IEEE Trans. Biomed. Eng.* BME-15, 169-179.
- de Boer, E. and de Jongh, H.R. (1978) On cochlear encoding: potentialities and limitations of the reverse correlation technique. *J. Acoust. Soc. Am.* 63, 115-135.
- Deng, L., Geisler, C.D. and Greenberg, S. (1987) Responses of auditory-nerve fibers to multi-tone complexes. *J. Acoust. Soc. Am.* 82, 1989-2000.
- Eckhorn, R. and Popel, B. (1979) Generation of Gaussian noise with improved quasi-white properties. *Biol. Cybernetics* 32, 243-248.
- Eggermont, J.J., Johannesma, P.I.M. and Aertsen, A.M.H.J. (1983) Reverse-correlation methods in auditory research. *Quart. Rev. Biophys.* 16, 341-414.
- Eggermont, J.J. and Smith, G.M. (1990) Characterizing auditory neurons using the Wigner and Rihacek distributions: a comparison. *J. Acoust. Soc. Am.* 87, 246-259.
- Epping, W.J.M. and Eggermont, J.J. (1985a) Single-unit characteristics in the auditory midbrain of the immobilized grassfrog. *Hear. Res.* 18, 223-243.
- Epping, W.J.M. and Eggermont, J.J. (1985b) Relation of bin-aural interaction and spectro-temporal characteristics in the auditory midbrain of the grassfrog. *Hear. Res.* 19, 15-28.
- Evans, E.F. and Nelson, P.G. (1973) The responses of single neurones in the cochlear nucleus of the cat as a function of their location and the anaesthetic state. *Exptl. Brain Res.* 17, 428-442.
- French, A.S. and Holden, A.V. (1971) Alias-free sampling of neuronal spike trains. *Kybernetik* 5, 165-171.
- Godfrey, D.A., Kiang, N.Y.-S. and Norris, B.E. (1975a) Single unit activity in the posteroventral cochlear nucleus of the cat. *J. Comp. Neurol.* 162, 247-268.

- Godfrey, D.A., Kiang, N.Y.-S. and Norris, B.E. (1975b) Single unit activity in the dorsal cochlear nucleus of the cat. *J. Comp. Neurol.* 162, 269–284.
- Hermes, D.J. (1985) Separation of time and frequency. *Biol. Cybern.* 52, 109–115.
- Hermes, D.J., Aersten, A.M.H.J., Johannesma, P.I.M. and Eggermont, J.J. (1981) Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (*Rana temporaria* L.) investigated with noise stimuli. *Hear. Res.* 5, 147–178.
- Johannesma, P., Aersten, A.D., Cranen, B. and Van Erning, L. (1981) The phonochrome: a coherent spectro-temporal representation of sound. *Hear. Res.* 5, 123–145.
- Kiang, N.Y.-S. (1965) *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve.* MIT Press, Cambridge.
- Kim, D.O., Sirianni, J.G. and Chang, S.O. (1990) Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: Analysis with autocorrelation/power-spectrum. *Hear. Res.* 45, 95–113.
- Møller, A. (1981) Latency in the ascending auditory pathway determined using continuous sounds: comparison between transient and envelope latency. *Brain Res.* 207, 184–188.
- Pfeiffer, R.R. (1966) Classification of response patterns of spike discharges for units in the cochlear nucleus: tone burst stimulation. *Exp. Brain Res.* 1, 220–235.
- Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986) *Numerical Recipes.* Cambridge: Cambridge Univ. Press.
- Rhode, W.S. and Smith, P.H. (1986a) Encoding timing and intensity in the ventral nucleus of the cat. *J. Neurophysiol.* 56, 261–286.
- Rhode, W.S. and Smith, P.H. (1986b) Physiological studies on neurons in the dorsal cochlear nucleus of cat. *J. Neurophysiol.* 56, 287–307.
- Rihaczek, A.W. (1968) Signal energy distribution in time and frequency. *IEEE Trans. on Inf. Th.* IT-14, 369–374.
- Sachs, M.B. and Kiang, N.Y.-S. (1968) Two-tone inhibition in auditory-nerve fibers. *J. Acoust. Soc. Am.* 43, 1120–1128.
- van Stokkum, I.H.M. (1987) Sensitivity of neurons in the dorsal medullary nucleus of the grassfrog to spectral and temporal characteristics of sound. *Hear. Res.* 29, 223–235.
- Swerup, C. (1978) On the choice of noise for the analysis of the peripheral auditory system. *Biol. Cybernetics* 29, 97–104.
- Young, E.D. and Brownell, W.E. (1976) Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats. *J. Neurophysiol.* 39, 282–300.