Linear and Nonlinear Spectral Integration in Type IV Neurons of the Dorsal Cochlear Nucleus. II. Predicting Responses With the Use of Nonlinear Models

ISRAEL NELKEN,1 PETER J. KIM,2 AND ERIC D. YOUNG2
1Department of Physiology, The Hebrew University—Hadassah Medical School, Jerusalem 91120, Israel; and
2Department of Biomedical Engineering and Center for Hearing Sciences, The Johns Hopkins University, Baltimore, Maryland 21205

Nelken, Israel, Peter J. Kim, and Eric D. Young. Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models. J. Neurophysiol. 78: 800–811, 1997. Two nonlinear modeling methods were used to characterize the input/output relationships of type IV units, which are one principal cell type in the dorsal cochlear nucleus (DCN). In both cases, the goal was to derive predictive models, i.e., models that could predict the responses to other stimuli. In one method, frequency integration was estimated from response maps derived from single tones and simultaneous pairs of tones presented over a range of frequencies. This model combined linear integration of energy across frequency and nonlinear interactions of energy at different frequencies. The model was used to predict responses to noisebands with varying width and center frequency. In almost all cases, predictions using two-tone interactions were better than linear predictions based on single-tone responses only. In about half the cases, reasonable quantitative fits were achieved. The fits were best for noisebands with narrow bandwidth and low sound levels. In the second nonlinear method, the spectrotemporal receptive field (STRF) was derived from responses to broadband stimuli. The STRF could account for some qualitative features of the responses to broadband noisebands and spectral notches embedded in broad noisebands. Quantitatively, however, the STRFs failed to predict the responses of type IV units even to simple broadband noise stimuli. For narrowband stimuli, the STRF failed to predict even qualitative features (such as excitatory and inhibitory frequency bands). The responses of DCN type IV units presumably result from interactions of two inhibitory sources, a strong one that is preferentially activated by narrowband stimuli and a weaker one that is preferentially activated by broadband stimuli. The results presented here suggest that the STRF measures effects related to the broadband inhibition, whereas two-tone interactions measure mostly effects related to narrowband inhibition. This explains why models based on two-tone interactions predict the responses to narrow noisebands much better than models based on STRFs. It is concluded that a minimal stimulus set for characterizing type IV units must contain both broadband and narrowband stimuli, because each stimulus class by itself activates only partially the integration mechanisms that shape the responses of type IV units. Similar conclusions are expected to hold in other parts of the auditory system: when characterizing a complex auditory unit, it is necessary to use a range of stimuli to ensure that all integration mechanisms are activated.

INTRODUCTION

The dorsal cochlear nucleus (DCN) in unanesthetized animals displays striking nonlinear response characteristics (Nelken and Young 1994, 1997; Spirou and Young 1991). As a result, it has been difficult to predict the responses of DCN principal cells to stimuli with arbitrary spectral shape. This paper focuses on type IV units, which are one response type recorded from DCN principal cells. Tones and narrow bands of noise evoke inhibitory responses in type IV units over a broad range of frequencies and sound levels (Spirou and Young 1991; Young and Brownell 1976); from these responses, one would predict that broadband noise (BBN) stimuli would also evoke inhibitory responses. However, the responses to BBN in type IV units are in fact predominantly excitatory. Other examples of nonlinear behavior, in which the response to the sum of two stimuli is far different from the sum of the responses to the two stimuli presented separately, have been documented (Nelken and Young 1997; Spirou and Young 1991).

To understand the nature of the representation of the acoustic surround in DCN, and in other similarly complex auditory structures, it is desirable to have general predictive models of the response characteristics of type IV units. That is, it is desirable to have a model that will predict the neurons’ responses to arbitrary acoustic stimuli. Such a model avoids the necessity of experimentally evaluating every new stimulus of interest and also provides a necessary basis for a comprehensive theoretical understanding of auditory signal processing in the brain.

Two approaches to such models have been taken. One approach is based on developing a physiologically based model that directly represents the neural interactions that shape the response properties of the structure. Qualitative (Nelken and Young 1994) and semiquantitative (Blum et al. 1995; Reed and Blum 1995) models of DCN have been developed that account for its known properties. However, this approach is not general and can be applied to a structure only after extensive experimental studies. The alternative is to develop models that capture the important aspects of the input/output processing in a structure, without requiring a detailed knowledge of its internal processing mechanisms (e.g., Marmarelis and Marmarelis 1978). A variety of quasilinear and nonlinear modeling techniques has been applied to the auditory system (e.g., Aertsen et al. 1981; Backoff and Clopton 1991; Eggermont et al. 1983b; Nelken et al. 1994a,b; Schreiner and Calhoun 1994; Shamma et al. 1995; Wickesberg et al. 1984). The predictive power of these
methods has been evaluated to a limited extent. In some cases, good predictions are obtained, e.g., predictions of responses to four-tone complexes from responses to single- and two-tone complexes (Nelken et al. 1994b) and predictions of responses to arbitrary spectral shapes from a linear superposition of rippled-spectra stimuli of different ripple frequencies (Shamma and Versnel 1995). However, other studies have emphasized poor results of prediction attempts, particularly when the stimuli used to derive the model differ substantially from the test stimuli (Bonham et al. 1996; Eggermont et al. 1983a).

The DCN is a promising test system for the development of nonlinear input/output modeling methods because of the detailed descriptions available of its internal synaptic organization and of the response properties of its inputs. In this paper, we consider the application of two nonlinear modeling methods to the DCN, one based on single- and two-tone response maps (Nelken et al. 1994a,b) and the other on the spectrotemporal receptive field (STRF) (Aertsen and Johannesma 1981). Both methods should capture some nonlinearities in the Wiener-Volterra sense. We emphasize predictions in which the levels of the stimuli used to develop the models and the stimuli used to test them are similar. Thus we emphasize the spectral integration capabilities of the models rather than their ability to accurately predict behavior as a function of level. For both methods, the crucial assumption for applicability is that the neglected higher-order terms are small relative to the terms that are kept in the model.

We show that the characterizations derived from these two methods differ, as do the predictive abilities of the two methods; each method is only able to predict responses to stimuli similar in spectral bandwidth to the stimuli used to develop the model. This result can be interpreted in terms of the current qualitative model of DCN synaptic organization, because the stimuli used in the two modeling methods activate different populations of inhibitory interneurons in DCN, and therefore different aspects of the nonlinear behavior of DCN type IV units.

METHODS

Animal preparation and protocol

The methods have been described in detail elsewhere (Nelken and Young 1997). Briefly, single-unit recordings were made in DCN of decerebrate cats with the use of tipped platinum-iridium microelectrodes. Responses to best-frequency (BF) tones and BBN rate functions were obtained to verify the stability of the unit. Because the design of the experiment required comparison of rate data obtained over periods of minutes to hours, unit stability was important. Recordings were not made from units unless spike isolation and triggering were clear and without artifact. Changes in unit properties over long periods of time still occurred in a minority of units (4 of 25 on which this report is based), but these occurred mostly for BBN at high levels, above those used for the comparisons in this paper. These units were therefore included in the sample analyzed here.

Responses to bands of noise of various bandwidths (Fig. 1A), to filtered noise (Fig. 1B), and to notch noise (Fig. 1C) were used to test the predictions of the modeling methods. The filtered noise was generated by filtering a pseudorandom BBN with a head-related transfer function (HRTF) for a cat (Rice et al. 1992). The HRTF is the transfer function from free field to a point near the eardrum; when a stimulus with this spectral shape is presented through a closed sound system, the result simulates free-field presentation of a BBN from a particular direction in space. The HRTF stimulus was stored in digital form and played through a D/A converter at a variable sampling rate chosen to place the prominent spectral notch in the stimulus (arrow) at various frequencies relative to unit BF.

The notch noise was generated by band-pass filtering (24 dB/octave) two independent BBNs with a low cutoff frequency of half the desired notch width and a high cutoff frequency equal to BF. The resulting noisebands were modulated in quadrature to the BF of the unit and added, generating a stationary noiseband with total bandwidth of 2 × BF and with a notch of the desired width ~30 dB deep centered on BF. Noisebands were generated similarly, by low-pass filtering (48 dB/octave) two independent BBNs with cutoff frequency equal to half the bandwidth and modulating the resulting bands to the desired center frequency. In later experiments, noise stimuli were generated digitally: the spectrum of the stimulus was generated in the frequency domain by filling the real and imaginary parts of the Fourier transform of the signal with independent Gaussian pseudorandom deviates (passbands) or zeros (stopbands). The resulting complex spectrum was then inverse transformed back into the time domain. The transform length was 32,768 points, giving a frequency resolution of 5 Hz at a sampling frequency of ~162 kHz. A 16-bit D/A converter was used to transform the digital samples into an analog waveform; stimulus level was set by attenuating the analog waveform after D/A con-
version. Because the Nyquist rate (>80 kHz) was above the audible range of cats, no antialiasing filter was used. Frequency bands generated in this way had much steeper slopes than the noisebands generated with the use of the analog modulation technique. Nevertheless, the responses to corresponding stimuli were found to be similar.

Linear and nonlinear summation of tones

The single- and two-tone response map data were used to predict responses to noisebands and notch-noise stimuli. Two models were used: a first-order model in which the response to a noise is predicted as the sum of the responses to tones within the passband of the noise, and a second-order model in which the summation is taken over all single- and two-tone combinations within the passband. Because the responses of DCN type IV units change dramatically with sound level, it is necessary to choose the level of the tones to be comparable with that of the noise; in this case, the tone level was set so that the tone power was equal to the power in a narrow noiseband of width $bw$ Hz, where the spectrum level of the narrow noiseband was the same as in the passband of the noiseband being predicted; $bw$ is called the reference bandwidth below and was usually 200 Hz.

For the first-order model, let $f_l$ and $f_u$ be the lower and upper edge of the noise passband to be predicted and $f_0 < f_1 < \cdots < f_n < f_{n+1}$ be the tone frequencies chosen such that

$$f_0 + f_1 \leq f_l$$

and

$$f_n \leq \frac{f_x + f_{x+1}}{2}$$

then the predicted response $R_{\text{noiseband}}^{1}$ was computed by

$$R_{\text{noiseband}}^{1} = R_{\text{spontaneous}} + \frac{1}{bw} \sum_{k=1}^{n} w_k (R(f_k) - R_{\text{spontaneous}}) \quad (1)$$

where $R_{\text{spontaneous}}$ is the spontaneous rate, $R(f_k)$ is the response to the tone of frequency $f_k$, and the weights $w_k$ account for the fact that the frequencies are usually logarithmically, rather than linearly, spaced

$$w_1 = \frac{f_1 + f_2}{2} - f_l \quad w_k = \frac{f_k - f_{k-1}}{2} \quad k = 2, 3, \ldots n - 1 \quad \text{and} \quad w_n = f_u - \frac{f_{n-1} + f_n}{2}$$

That is, each $R(f_k)$ is assumed to represent the response rate to noise energy in a frequency band of width $w_k$ centered on $f_k$. The factor $w_k/bw$ in Eq. 1 corrects the tone response rate $R(f_k)$ for the fact that the tone is used to approximate a bandwidth $w_k$ of the noise signal, not the reference bandwidth $bw$, for which the level of the tone is appropriate. Note that we assume that it is driven rates that summate linearly, not total rates; that is, spontaneous is the spontaneous rate, $D^1$ and $C^2$ are the rates to the tones individually, as above. The spontaneous rate is added to compensate for the fact that it is subtracted twice in the $R(f_k)$. Note that $c(f_1, f_2) = 0$ if the two tones combine linearly in the sense of Eq. 1. The nonlinear terms $c(f_1, f_2)$ were computed from the single- and two-tone data with the use of Eq. 2. The second-order prediction $R_{\text{noiseband}}^{2}$ was obtained from $R_{\text{noiseband}}^{1}$ by adding terms derived from $c(f_1, f_2)$ as

$$R_{\text{noiseband}}^{2} = R_{\text{noiseband}}^{1} + \frac{bw}{f_u - f_l} D^1 + \frac{bw}{f_u - f_l} C^2 \quad (3)$$

where

$$D^1 = \frac{1}{2bw} \sum_{k=1}^{n} w_k c(f_k, f_1)$$

$$C^2 = \frac{1}{2bw^2} \sum_{k=1}^{n} w_k c(f_k, f_2)$$

The summation is taken over the region $f_1 < f_k < f_u$ and $f_k < f_1 < f_u$ and the weighting factor $w_k$ is defined similarly to $w_k$ in $R_{\text{noiseband}}^{1}$ and $D^1$. The need for the diagonal term $D^1$ is explained in the APPENDIX.

The factor $bw/(f_u - f_l)$ in Eq. 3 was developed empirically. An intuitive explanation for this correction factor is as follows: on the
one hand the neuron sums up roughly $n$ inputs where $n$ is proportional to the integration bandwidth of the neuron; on the other hand, the summation defining $C^2$ has contributions from about $n^2$ two-tone contributions. The correction factor $bw(f_1 - f_2)$ is roughly proportional to $1/n$, because it measures the width of the noiseband in units of $bw$ and so fixes this discrepancy.

The theory presented above can be extended to predict the responses to noisebands as function of their level. The second-order prediction in this case is

$$R_{\text{noiseband}}(L) = R_{\text{spontaneous}} + \frac{L}{L_{\text{ref}}} \left( R_{\text{noiseband}}^{1} - \frac{bw}{f_1 - f_2} D_1 - R_{\text{spontaneous}} \right)$$

$$+ \left( \frac{L}{L_{\text{ref}}} \right)^2 \frac{bw}{f_1 - f_2} C^2$$ (4)

Where $L$ is the level of the noiseband, $L_{\text{ref}}$ is the reference level (at which the 2-tone data were presented), and the other terms are those computed for the prediction of the noiseband at the reference level. The dependence of the various first- and second-order terms on the level is derived from Eq. A1. The first-order terms ($R_{\text{noiseband}}$ and $D_1$) depend only on the level at a single frequency, and therefore scale linearly with level. The second-order term ($C^2$) depends on products of tone levels at two frequencies, and therefore scales quadratically with level. Note that for $L = L_{\text{ref}}$, Eq. 4 reduces to Eq. 3. These rate-level predictions were compared with the measured rate-level functions, which were always taken for noisebands centered at the BF.

Note that although the terms ‘‘first- and second-order predictions’’ are used here, it is the power spectrum of the stimulus that is used as the input to the prediction formulas. These predictions are therefore actually second order and fourth order when considered as functions of the stimulus waveform. It follows that the predictions based on the STRF are actually comparable with the first-order predictions with the use of the formulation above (Eq. 1), although they are usually considered to be second order (in terms of the stimulus waveform).

**STRF calculations**

STRFs were computed with the use of the same algorithms as in Kim and Young (1994). An STRF is a function of time and frequency that measures the average spectral density of the stimulus as a function of time before action potentials from a neuron. A peri-stimulus time histogram was computed from the responses to the pseudorandom BBN. For each bin of the peri-stimulus time histogram, the time-frequency distribution of the sound segment (10–20 ms) just preceding it was computed (the Wigner distribution was used here) (Eggermont and Smith 1990; Kim and Young 1994), and those distributions were averaged over all bins with weights proportional to the number of spikes per bin of the peri-stimulus time histogram. This is equivalent to averaging the Wigner distribution of the sound preceding each spike, as in Eq. 7 of Kim and Young (1994). The expected time-frequency distribution in the absence of correlation between spikes and sound was subtracted to account for small spectral irregularities in the noise; the result of the subtraction is the STRF. The expected distribution was computed as the STRF for a sequence of Poisson spikes, uncorrelated with the pseudorandom BBN. The STRF can be thought of as representing the average spectrotemporal “triggering event” for the unit and for the stimulus used to compute it (Aertsen et al. 1981).

**RESULTS**

This paper is based on the responses of 25 type IV units to the two-tone stimulus paradigm. The responses of these units to other stimuli are described in the companion paper, which treats regions of linearity in type IV responses (Nelken and Young 1997). The STRFs of nine type IV units were collected; three of these units were also tested with two-tone stimuli and noisebands. The results are presented in two stages. First it will be shown that two-tone stimuli can be used to model some nonlinear phenomena, mostly for narrowband stimuli. Second it will be shown that the STRF predicts some aspects of responses to broadband, but not narrowband, stimuli.

**Two-tone stimuli can be used to predict the responses to narrow noisebands**

Examples of two-tone response planes for three different units are shown in Fig. 2. The axes of each plane are the two frequencies composing the stimulus. The firing rate is represented by gray levels, according to the scale to the right of each plane. The line plot below each plane shows the responses to single tones on the same frequency axis. Figure 2A shows a two-tone response plane for one unit at a low level, close to the maximum of the BF rate-level function. Significant responses occur only when the two-tone combination contains near-BF tones, resulting in the cross pattern in the figure. In Fig. 2A, the responses along the diagonal (where the 2 tones are identical and behave as a single tone with a level 6 dB higher) are almost the same as the single-tone responses off the diagonal, because the former is measured at a level slightly above and the latter slightly below the maximum of the BF rate-level function. The nonlinear interaction term $c(f_1, f_2)$ (Fig. 2D) is essentially zero except when both tones are very close to BF, where it is negative.

Figure 2B shows, for a different unit, a two-tone response plane at a higher level, on the descending limb of the BF rate-level function. In this case, the responses on the diagonal are noticeably smaller than the single-tone responses, because of the nonlinear rate-level function. In this case, $c(f_1, f_2)$ (Fig. 2E) shows a larger range with significant interactions. Figure 2C shows a two-tone response plane at still higher levels for a third unit. Here the single-tone response is inhibitory at most frequencies. Note, however, that the two-tone responses (in Fig. 2F) show definite facilitation, for example the white spots near frequencies of 6.13 and 6.88 kHz, neither of which is excitatory when presented by itself. The two-tone response planes at low and medium levels invariably had the shape of a cross, showing that two-tone interactions occur mainly between frequency components that are close to each other and to BF (Fig. 2, D and E). At high levels, however, the shape of the two-tone response plane was highly variable (Fig. 2F).

First- and second-order predictions for the responses to narrow noisebands (Fig. 1A) were generated from the tone responses with the use of Eqs. 1 and 3. The noisebands had fixed bandwidth and spectral level, but varied in center frequency. The plots in Fig. 3 show discharge rate versus center frequency. The examples in Fig. 3 cover the range of goodness of fit that can be achieved with single- and two-tone models. In Fig. 3A, the first-order fit (•••) is generally too large. The second-order fit (−−−) follows closely the measured response (-----), except for some oscillations at center frequencies between 4 and 5 kHz. This oscillation...
Significant but weak response over a large frequency range. The first-order prediction oscillates wildly above and below the measured response. The second-order correction often has the right sign, but is too large at most frequencies. This creates the same pattern as in Fig. 3D: for frequencies at which the linear prediction is too low, the second-order prediction tends to be too large and vice versa.

The quality of the fits of the measured and predicted data were quantified with the use of an objective measure $d$ equivalent to the $\chi^2$ of the difference between the measured and predicted functions; if the difference between two functions is entirely due to noise, then $d$ should be near 1, and $d$ increases as the difference between the functions increases. This measure is described in detail elsewhere (Nelken and Young 1997). In the predictions made here, cases with $d \leq 1$ were rare. This is probably the result of two factors: first, the different character of the predicting stimuli (2-tone complexes) and the predicted stimuli (noisebands); and second, may be the result of noise in the measurements. Figure 3B is another example of a good fit. Figure 3C shows a borderline case. Although the second-order prediction correctly identifies the main peak of the response and its bandwidth, the prediction for the maximal rate is too large by a factor of $\sim 2$.

Figure 3D, E shows examples of bad fits. In Fig. 3D, the first-order prediction is too small most of the time. It is also qualitatively wrong, showing a dip where the actual response has a peak. The second-order correction fixes the qualitative discrepancy, but the correction is too large. As a result, the second-order prediction does show a peak at about the right frequency, but the rate at the peak is much too large. Finally, Fig. 3E shows a case in which there was a significant but weak response over a large frequency range. The first-order prediction oscillates wildly above and below the measured response. The second-order correction often has the right sign, but is too large at most frequencies. This creates the same pattern as in Fig. 3D: for frequencies at which the linear prediction is too low, the second-order prediction tends to be too large and vice versa.

The quality of the fits of the measured and predicted data were quantified with the use of an objective measure $d$ equivalent to the $\chi^2$ of the difference between the measured and predicted functions; if the difference between two functions is entirely due to noise, then $d$ should be near 1, and $d$ increases as the difference between the functions increases. This measure is described in detail elsewhere (Nelken and Young 1997). In the predictions made here, cases with $d \leq 1$ were rare. This is probably the result of two factors: first, the different character of the predicting stimuli (2-tone complexes) and the predicted stimuli (noisebands); and second,
the noise in the estimates of the second-order contributions (they are computed as the difference between 3 measured rates, increasing their SD by a factor of ~1.7). Fits with \(d < 30\) were clearly good, as judged subjectively; fits with \(d > 300\) were clearly bad. In between, there was a large group of results for which the objective \(d\) value and the subjective judgment did not correlate well (e.g., Fig. 3, C vs. D, both of which have similar \(d\) values). Values of \(d\) for the second-order fits are shown in Fig. 3, A–E.

A total of 73 two-tone response maps was recorded from the 25 type IV units. From these, 428 predictions were made. Fits with \(d \sim 1\) were found for 18.5% of the cases (\(d < 2: 79\) of 428). About half the cases had \(d < 30\), showing reasonable fits (242 of 428, 57%). Another third of the cases lay in the middle range, \(30 < d < 300\), where correlation between \(d\) and subjective fit quality was weak (128 of 428, 30%). Finally, 13% of the cases had large \(d\) values (\(d > 300: 58\) of 428).

Figure 4A shows a scatter plot of \(d\) against the bandwidth of the test noiseband; the bandwidth is measured in octaves relative to the reference noise bandwidth used in the prediction (\(bw\) in Eqs. 1 and 3, usually 200 Hz). The plot shows a tendency for the fit to degrade with bandwidth. In Fig. 4B, the same data are shown, but now bandwidth is measured relative to the BF of the unit. Now the plot is wedge shaped, showing larger scatter when fitting bandwidths that are comparable with BF. That is, the fit can be good or bad if the test noiseband is wide compared with BF, but the fit is always bad if the test bandwidth is wide compared with the reference bandwidth.

Figure 4C shows a scatter plot of \(d\) values for predicting the responses to the same noise bandwidth at two different levels. There is a correlation between the fits at the two levels: units with bad fits at low levels tended to show bad fits at higher levels also. The line through the scatter is the equality line. In the majority of the cases, the fit at the higher level was worse than the fit at the lower level (\(P = 0.0025, 1\)-tailed Wilcoxon test).

An unresolved question in nonlinear modeling of auditory neurons is whether the second-order kernels actually improve prediction quality. For example, Wickesberg et al. (1984) showed that for units in the anteroventral cochlear nucleus, the second-order Wiener kernels did not improve response predictions in some cases. For the method presented here, the second-order prediction was almost always better than the first-order prediction, as judged by the \(d\) measure. In Fig. 4D, the scatter plot of the \(d\) values for first- and second-order predictions lies almost entirely below the diagonal. The \(d\) value in some cases decreased by 3 orders of magnitude.

In addition to predicting the responses to bands at a fixed level with varying width and center frequency, it is possible to predict, with the use of similar formulas, the responses of type IV units to noisebands of varying level (Eq. 4). Such predictions are shown in Fig. 5. Figure 5, A and B, shows cases in which the second-order formula, based on response plots obtained at one sound level, actually predicts the nonmonotonicity of the type IV units. Figure 5C shows a case in which the second-order prediction is monotonic over the relevant level range, but the fit at low levels (below the peak of the rate-level function) is excellent. Finally, Fig. 5D shows a case in which the prediction was nonmonotonic, but the quantitative fit is bad—threshold is too low and the maximum firing rate of the unit is too high. As expected, there was strong correlation between cases in which the fixed-level predictions (i.e., Fig. 3) were good and cases in which the rate-level predictions were good. Twenty-one units were tested with at least one noiseband rate-level function; for a large majority (19 of 21) the rate-level prediction was successful (at least at low levels, as in Fig. 5C) at some bandwidths. Nonmonotonicity was predicted in 16 of 21 of
the units. Only cases in which the two-tone data were measured close to the peak of the rate-level function (as in Fig. 2, A and B) showed reasonable predictions for noiseband rate-level functions, and the predictions were best for narrower noisebands (200–800 Hz). These results show that, in some circumstances, the two-tone response maps capture not only information about spectral integration at a fixed level but also the information needed to describe the behavior of type IV units as a function of level.

**STRFs of DCN type IV units**

STRFs were computed for nine type IV units, usually at multiple levels. Figure 6 shows three examples of type IV STRFs. The STRF is interpreted as the mean triggering event for spikes, in the sense that the STRF is an estimate of the average power spectrum in the stimulus as a function of time preceding spikes. A triggering event may be either an increase in the level of some frequencies in the noise (white regions in Fig. 6) or a decrease in the level of other frequencies (black regions in Fig. 6), or both. Spike latency is the time from the triggering event to the spike, which occurs at time 0 on the ordinate of the STRF plots.

Figure 6A shows a unit that had a strong, narrow excitatory region at a latency of 3–4 ms. It was surrounded on three sides (that is, at higher and lower frequencies and on BF at longer latency) by an inhibitory region. This is the most commonly seen pattern of STRF. A similar arrangement, although weaker, appears again at longer latencies.

It is suspected that the second, longer-latency repeat of the excitatory/inhibitory regions in Fig. 6A is in fact an artifact due to regularities in the spike discharge of the type IV unit (Backoff and Clpton 1991; Kipke et al. 1991; Parrham and Kim 1992). To check this possibility, the autocorrelation was computed for this unit; the interval between the two peaks in Fig. 6A (~1.5 ms) was within the refractory period, so the second peak is not produced by simple regularity in this unit. However, multipeaked STRFs were observed in three other units and, in those cases, the autocorrelation showed a significant peak at the repeat period of the STRF. Thus, in the majority of cases, repetitions in the STRF result from regularity of firing of the unit; that is, they are related to the intrinsic properties of the unit rather than to spectral integration mechanisms. A possible mechanism to explain the exceptional case in Fig. 6A is that the unit has an intrinsic oscillatory rhythm at the right frequency (~600 Hz), which is not observed ordinarily in spike discharges because of refractoriness. However, in those cases in which a spike was not discharged at the usual short latency (~4 ms) from a favorable triggering event, the triggering event still evoked the intrinsic oscillation, which increased the firing probability 1.5 ms later.

In general, the STRF shows very strong correlation along the temporal dimension, which causes vertical stripes to appear. The vertical stripes result from the use of a finite-length noise segment to compute the STRF. They correspond to the peaks and valleys of the nonflat power spectrum of the specific noise segments used. Although some of this effect is reduced by subtracting the expected STRF under the no-correlation condition, obviously the compensation is not full.

Another configuration of excitatory and inhibitory regions is shown in Fig. 6B. Here at suprathreshold levels the excitatory region is very wide, extending from ~4.5 to 7.5 kHz. It also has a weaker excitation at lower and higher frequencies, extending essentially across the whole frequency axis of the STRF. At low frequencies, the latency of the weak excitatory region increases slightly as frequency decreases, which may be a correlate of the traveling wave in the cochlea. It might be hypothesized that such a wide excitatory region would
be manifested in the tone responses of the unit. Figure 7A shows the single-tone response map of this unit. This plot shows discharge rate as a function of frequency at eight sound levels. The horizontal lines are spontaneous rate; inhibitory regions, where rate is less than spontaneous, are shaded and excitatory regions are shaded in black. The unit does indeed have an excitatory region near threshold (−80 dB), which has a width similar to that of the strong excitatory region in the STRF; however, above 10 dB re threshold, the unit gave predominantly inhibitory responses to tones. Thus there is no qualitative correspondence between the tone response map and the STRF.

Figure 6C shows a third configuration of excitation and inhibition, which is to some extent intermediate between the first two. There is a relatively wide short-latency excitatory region that seems to be restricted at longer latencies by inhibitory regions on either side of the BF.

The STRFs are a sensitive measure of DCN units' response characteristics in that STRFs vary considerably from unit to unit but, in most units, their shapes stay relatively constant as sound level changes. Figure 7 shows a comparison of tone response maps (Fig. 7, A and C) with STRFs (Fig. 7, B and D) for two units. The STRFs are shown as frequency marginals, which are computed by averaging the STRFs along the time axis over a range of latencies; they show the frequency tuning in the STRF, with variations in time averaged out. The solid lines show frequency marginals computed over latencies that include the principal excitatory peak, but not the inhibitory valleys at longer latency. The dashed lines show frequency marginals computed over latencies that cover the whole response area, including both the excitatory peak and the subsequent inhibitory valleys. The dashed lines therefore show the average deviation of the STRF from background across all latency values. The example shown in Fig. 7B is from the same unit as Fig. 6B. The broad excitatory area is clear in the solid lines; the dashed lines show that the inhibitory input centered at 5.7 kHz has a strong effect on the cell, producing a narrow net inhibitory input just below BF at high levels. The example in Fig. 7D is more typical of the type IV units in our sample in that it has an excitatory region centered at BF, an inhibitory region above BF, and a general shape that does not change with stimulus level. All STRFs had excitatory regions centered on BF, except for one case, where inhibitory regions only were seen at high levels (−20 dB above threshold).

Most type IV STRFs (7 of 9) showed significant inhibitory regions. Inhibitory regions at latencies just longer than excitatory regions (e.g., Fig. 6, A and B) can sometimes be attributed to refractoriness (Kim and Young 1994). Inhibitory regions at frequencies just above or just below BF were also observed in most units (6 of 8); the cases in Figs. 6B and 7B is one exception. The cases in Figs. 6, A and C, and 7D are more typical. These off-BF inhibitory regions probably result from neural interactions. In four of six cases with both BF excitation and off-BF inhibition, the inhibitory region had a slightly longer latency than the excitatory region, which is consistent with a multisynaptic inhibitory pathway. Very small off-BF inhibitory regions are observed in auditory nerve fibers (Kim and Young 1994), but these are too small to explain the kind of inhibitory effects shown in Fig. 6.

Prediction of responses to other stimuli based on the STRF

The STRF has limited predictive power in the sense that it cannot be used to describe the responses of DCN units to most other stimuli. It can be seen in Fig. 7 that the STRF fails to predict the responses of most type IV units to tones. In fact, except for the case in Fig. 7B and one other in which the STRF was entirely inhibitory, the STRF predicts excitatory effects of BF tones at all levels; this occurs because the largest component of type IV STRFs at all levels is the excitatory peak near BF. By contrast, type IV units show inhibitory responses to BF tones at suprathreshold levels. This failure of the STRF to predict inhibitory responses
to tones extends also to the prediction of the responses to narrow noisebands, because they give essentially the same responses as do tones (Nelken and Young 1997).

It can be shown theoretically that the STRF can be used to predict rate responses to stimuli with arbitrary spectra by multiplying the frequency marginal of the STRF by the spectrum of the stimulus and summing across frequency. The ability of STRFs to predict the responses to broadband stimuli was tested with the use of this calculation. STRF predictions were compared only qualitatively with unit responses, because shifting and scaling were necessary to make the ranges of the predictions approximately equal to those of the measured rates. Four units were tested with BBN filtered to have the spectrum of a cat HRTF (Fig. 1B). In these cases, the bandwidth of the excitatory peak in the STRF was a good but qualitative predictor of the sensitivity of the unit to the spectral notch. The unit in Fig. 6B was rather insensitive to the center frequency of the notch, whereas units with a narrow peak in their STRF (as in Fig. 6A) were much more sensitive to the frequency of the notch relative to their BFs. For two additional units, the STRF failed to predict responses to spectral notches embedded in BBN (Fig. 1C). These units were unusual for DCN type IV units in that they were inhibited by BBN. The STRF failed in these cases by predicting an excitatory response to the noise. Thus the STRF has limited ability as a predictor of responses to wideband stimuli.

**DISCUSSION**

**Nonlinear modeling methods**

The arsenal of nonlinear modeling methods is rather limited at present. The most widely tested methods for nonlinear modeling of auditory units have been second-order reverse-correlation-based methods, used either in the time or frequency domains (Wiener kernel) (Wickesberg et al. 1984) or as the STRF (Aertsen and Johannesma 1981; Backoff and Clopton 1991; Eggermont et al. 1983a). In all cases, some function of the waveform just preceding a spike is averaged over all spike occurrences. Various forms of STRF have been used, but all are related to the Fourier transform of the second-order Wiener kernel across one time dimension, giving the average time-frequency distribution of energy preceding a spike.

The usefulness of the STRF for predicting responses has been found to be limited (Eggermont et al. 1983a). There are two problems: first, the STRF assumes a stationary system and is generated from steady-state stimuli, whereas the responses of auditory neurons to interesting stimuli are not stationary. That problem is bypassed in this paper by at-
tempting to predict only average discharge rate behavior and not detailed temporal aspects of responses. This is appropriate in DCN, because average discharge rate captures reasonably well the temporal modulation of units' responses (Young and Brownell 1976). The second problem is that there is no reason to assume a priori that the nonlinearity of an auditory neuron is limited to second-order terms in the Wiener-Volterra sense. As is discussed below, the results of this paper suggest that the nonlinearities are of higher than second order in the DCN. Therefore, although the full Wiener-Volterra series would represent the unit completely, there is no reason to think that truncating it at the second order would give a useful approximation (Johnson 1980). Unfortunately, estimating higher-order kernels is technically difficult.

The alternative approach to characterizing nonlinear neural systems is to build physiological models based on the internal organization of the system. In the case of the DCN, such models implement some or all of the known anatomic and physiological facts about DCN type IV units, and then test their ability to predict the responses to various stimuli. Three such models have been suggested in the past (Arle and Kim 1991; Blum et al. 1995; Pont and Damper 1991; Reed and Blum 1995). Although these models show some promise, they have not yet demonstrated predictive power for the kinds of stimuli analyzed here.

Tone and two-tone predictions

First-order predictions of responses to noisebands with the use of summation of responses to single tones have been found to have very limited power (Spirou and Young 1991). In this paper, we show that two-tone responses can be used to improve the quality of these predictions, but that the improvement depends on the bandwidth of the noise and also on its level (Fig. 4). The second-order predictions almost always improve the fit qualitatively, but sometimes overcompensate for the errors in the single-tone fits (e.g., Fig. 3D). Presumably, this behavior reflects the fact that higher-order nonlinear terms are necessary to fully model the unit's responses. In addition, there are indications that some of the problems in the second-order prediction are caused by noise in the two-tone responses, because only one repetition of each combination of parameters was presented and the two-tone contributions are computed as differences of these noisy values.

In the companion paper (Nelken and Young 1997), we argue that the major source of nonlinearity in type IV responses is the inhibitory input from type II units. The limited success that was achieved with second-order predictions is probably due to the fact that type II units are relatively weakly activated in the parameter range used. This follows from the fact that the tone levels used for the single- and two-tone response maps were usually at the peak of the rate-level function or on its descending limb, which is where type II units are just beginning to fire (Young and Voigt 1981). Assuming that type II units are the major source of nonlinearity, the fact that the predictions deteriorated with increasing bandwidth (Fig. 4A) can be explained as resulting from an increasing difference between the narrowband predictor, which activates the type II units, and the broadband test response, which does not.

The ability of the second-order predictions to capture the nonmonotonicity of the rate-level function of narrow noisebands (Fig. 5) may be explained with a similar argument. Because the nonlinearity introduced by the type II unit is weak, it can be approximated by parabolic term, as in Eq. 4. At higher levels the rate-level function is no longer parabolic—in fact, the units usually shut down completely. As a result, the predictions based on two-tone measurements cannot predict the rate-level functions at higher levels. It is nevertheless noteworthy that one measurement, near the peak of the rate-level function, is sometimes able to capture the quantitative features of the whole rate-level functions from threshold to peak and to the inhibitory area.

**STRF predictions**

In contrast to the two-tone predictions, the STRFs of type IV units are computed with the use of responses to broadband stimuli. They do show qualitative agreement with the responses of some type IV units to notch stimuli, but are unable to predict the responses to narrowband stimuli (Fig. 7). In this sense, they complement the two-tone data, which are taken with the use of narrowband stimuli and are useful for predicting the responses to some narrowband stimuli. The main use of the STRF may be in defining and classifying the large variability in properties of type IV units (Nelken and Young 1994; Spirou and Young 1991; Young and Brownell 1976).

The STRF results contain one puzzle: the extent of the inhibitory regions seen in the STRFs is much narrower than is expected from direct measurements of inhibitory bandwidths with broadband notch-noise stimuli, where the inhibitory bandwidth can be up to 1 octave re BF (median value ~0.5 octave) (Nelken and Young 1994). With broadband stimuli, type II units are minimally activated and the inhibition observed with notch-noise stimuli probably comes from a second inhibitory interneuron, the wideband inhibitor. The wideband inhibitor’s inhibition of type IV units is hypothesized to be weak (Nelken and Young 1994), and the STRF suffers from a narrow dynamic range (Kim and Young 1994). Thus the explanation for the difference in inhibitory bandwidth may be that only the central, strongest portion of the wideband inhibitor’s input is reflected in the STRF. Alternatively, the narrow inhibitory inputs in the STRF may be the weak remains of type II inhibition or of inhibition from other unknown sources.

The STRF data also support the conclusion that the nonlinearity of type IV units is of higher than second order in the Wiener-Volterra sense. Although the general position and extent of excitatory and inhibitory domains in the STRF are consistent with the responses to broadband stimuli, quantitative predictions fail. Higher-order kernels are therefore needed to achieve quantitative agreement with the data, even for broadband stimuli.

**Implications for auditory modeling**

The main advantage of the modeling approach used in this study is that there are no free parameters in the prediction formulas, and therefore no training. The formulas are completely general, and they are made specific to DCN type IV
units by plugging in the measured responses to a selected
family of sounds. However, the results presented here show
that it is very important to select correctly a family of sounds
that captures the whole complexity of the integration mecha-
nisms of the unit.

In this paper, two families of sounds were used: two-
tone stimuli and BBN. Although this made it possible to
characterize the responses of type IV units in some parameter
ranges, these stimuli turned out to be insufficient for a full
description of the responses of type IV units. One probable
cause for this failure is that neither family activated both of
the inhibitory sources on type IV units at the same time.
Therefore the responses to sounds that activate both sources
were not predicted correctly.

A possible set of sounds that might cover better the rele-
vant parameter regimes would be composed of noisebands
and pairs of noisebands of various bandwidths and separa-
tions. At the limiting case of zero bandwidth, pure tones
would be used. This set is four-dimensional (2 coordinates
of frequency and 2 coordinates of bandwidth) and therefore
is difficult to characterize with fine sampling of all four
coordinates. A judicious choice of bandwidths, with interpo-
lation in between, must be made to make the use of this set
practical.

This paper uses formal modeling techniques similar in
spirit to the Wiener-Volterra approach. It does, however,
spread a wider net, by using a large number of stimuli and
testing essentially all the possible interrelations between
them. The choice of the stimuli is based on previous physi-
ological knowledge. It is hoped that in this way the nonlinear
behavior can be explained better and parameter regimes that
are still not well understood be identified. The results may
be used in this way to direct further experiments and suggest
new families of sounds that may be used to study DCN
spectral integration mechanisms. Ultimately, of course,
physiological models will be needed to provide complete
explanation of DCN responses.

APPENDIX

In this appendix, mathematical derivations related to the prediction
formulas are outlined. The basic model used here is the follow-

\[ R[I(f)] = \int K_1(f)I(f)df + \frac{1}{2} \int \int K_2(f_1, f_2)I(f_1)I(f_2)df_1df_2 + R_{\text{spontaneous}} \]

(A1)

Where \( I(f) \) is the stimulus spectrum, \( R[I(f)] \) is the predicted rate
in response to the stimulus with spectrum \( I(f) \), \( K_1 \) is called the
first-order kernel, and \( K_2 \) is called the second-order kernel.

For a first-order model, in which the second-order kernel is null,
the single tones are sufficient. Single tones are delta functions in the
frequency domain, and so

\[ I(f) = \delta(f-f_1) \]

\[ R[I(f)] = K_1(f_1) + R_{\text{spontaneous}} \]

(A2)

Note that we assume here that the units of the tone level are fixed
so that the level used is equal to 1. In this case, the kernel \( K_1(f) \)
is just the single-tone response \( R(f) \) minus the spontaneous rate;
by approximating the integral in Eq. A1 with sums and adding the
frequency weighting factors, one gets Eq. 1 in the text.

For the second-order model, in which both kernels are assumed
to be nonzero, both single-tone and two-tone data are necessary.
The following combinations are used

\[ I(f) = \delta(f-f_1) \]

\[ I(f) = 2 \delta(f-f_1) \]

\[ I(f) = \delta(f-f_1) + \delta(f-f_2), f_1 \neq f_2 \]

Here the first combination describe single tones at the nominal
level, the second describes single tones that are 6 dB louder (these
will be the tones on the diagonal of the 2-tone response map), and
the third describes combinations of different frequencies.

The predicted responses to these stimuli are, respectively

\[ R(f_1) = K_1(f_1) + \frac{1}{2} K_2(f_1, f_1) + R_{\text{spontaneous}} \]

\[ R(f_1, f_1) = 2 K_1(f_1) + 2 K_2(f_1, f_1) + R_{\text{spontaneous}} \]

\[ R(f_1, f_2) = K_1(f_1) + K_2(f_1, f_2) + \frac{1}{2} K_2(f_1, f_2) + \frac{1}{2} K_2(f_2, f_1) + R_{\text{spontaneous}} \]

From these three equations, it is possible to extract the kernels

\[ K_1(f_1) = R(f_1) - \frac{1}{2} K_2(f_1, f_1) - R_{\text{spontaneous}} \]

\[ K_2(f_1, f_2) = R(f_1, f_2) - R(f_1) - R(f_2) + R_{\text{spontaneous}} \]

(A3)

In the text (Eq. 2), \( K_1(f_1) \) is denoted \( c(f_1, f_1) \). Note that the
first-order kernel \( K_1(f) \) is different in the first- and second-order
models (compare Eqs. A2 and A3); for the second-order model,
the first-order kernel is corrected by half the diagonal of the second-
order kernel. In Eqs. 3 and 4 in the text, this is expressed by
correcting the first-order prediction \( R \) with the term \( D \). It can be
shown that this term exactly cancels the contribution from the
diagonal in the double integral that defines the second-order term.
In principle, the prediction could be performed without correcting
the first-order kernel, provided that the diagonal was set to zero in
the double integral. However, this approach would not work when
predicting the responses to noisebands of varying level (as in Eq.
4). In this case, the corrected first-order kernel is required.

In practice, it turned out that the second-order correction is
very often in the right direction but too large. The discrepancy
increased with bandwidth, and it turned out that over the popula-
tion, the increase in discrepancy was proportional to the band-
width being predicted. This finding led to the introduction of the
factor \( bwL(f_1-f_2) \) to scale the second-order correction. The sur-
prising finding is that the proportionality factor is not significantly
different from 1, although there is no a priori reason for that. In
the physics literature, it is known that infinite series may sometimes
be summed by multiplying a lower-order term in the series by a
scaling factor. Further investigation of this finding may bring
additional insight into the spectral integration mechanisms of type
IV units.

The comments of Prof. Haim Sompolinsky were helpful in preparing this
manuscript.

This work was supported by National Institute of Deafness and Other
Communications Disorders Grant DC-00115 and by a grant from the Israeli
Academy of Sciences.

Address for reprint requests: I. Nelken, Dept. of Physiology, Hebrew
University—Hadassah Medical School, PO Box 12272, Jerusalem 91120,
Israel.

Received 30 August 1996; accepted in final form 15 April 1997.
REFERENCES


