Effects of Stimulus Spectral Contrast on Receptive Fields of Dorsal Cochlear Nucleus Neurons

Lina A. J. Reiss, Sharba Bandyopadhyay, and Eric D. Young

Center for Hearing and Balance, Johns Hopkins University, Baltimore, Maryland

Submitted 27 November 2006; accepted in final form 30 July 2007

Reiss LAJ, Bandyopadhyay S, Young ED. Effects of stimulus spectral contrast on receptive fields of dorsal cochlear nucleus neurons. J Neurophysiol 98: 2133–2143, 2007. First published August 1, 2007; doi:10.1152/jn.01239.2006. Neurons in the dorsal cochlear nucleus (DCN) exhibit strong nonlinearities in spectral processing. Low-order models that transform the stimulus spectrum into discharge rate using a combination of first- and second-order weighting of the spectrum (quadratic models) usually fail to predict responses to novel stimuli for principal neurons in the DCN, even though they work well in ventral cochlear nucleus. Here we investigate the effects of spectral contrast on the performance of such models. Typically, the models fail for stimuli with natural-sound–like spectral contrasts (~12 dB), but have good prediction performance at small (3-DB) contrasts. The weights also typically increase substantially in amplitude at smaller spectral contrast. These changes in weight size with contrast are partly inherited from similar effects seen in auditory nerve fibers, but there must be additional effects from inhibitory circuits in the DCN. These results provide insight into the reasons for the poor performance of spectrotemporal receptive field (STRF) models in predicting responses of auditory neurons. Because the general shapes of the weights do not change between low and high contrast, they also suggest that STRFs may capture meaningful properties of neural receptive fields, even though they do not do well at predicting responses.

INTRODUCTION

The receptive field of an auditory neuron shows how the neuron integrates its inputs across stimulus frequency and time. Often, receptive fields are defined on the basis of the assumption that discharge rate is a weighted linear sum of the components of the stimulus spectrum at different latencies. The parameters of such first-order models are estimated with reverse-correlation or similar methods, giving the neuron’s spectrotemporal receptive field (STRF; Aertsen and Johannesma 1981; Eggermont et al. 1983b; Escabi and Read 2003; Klein et al. 2000; Theunissen et al. 2000). These models show the selectivity of the neuron in a two-dimensional display with frequency on one axis and time on the other. Along the frequency axis, the STRF shows the neuron’s frequency selectivity or tuning; along the time axis, the STRF gives information about the latency of the neuron and temporal delays between responses to different frequency components.

The STRF provides a robust and easily computed model of the receptive field of a neuron. However, it is useful to know to what extent it is a complete model or a unique one, especially because it cannot capture all the features of a strongly nonlinear neuron (Johnson 1980). A rigorous test of the overall quality of such a model is its ability to predict responses to stimuli not used in the construction of the model. First-order models like the STRF work well in this test for peripheral neurons, such as auditory nerve (AN) fibers or neurons in the ventral cochlear nucleus (VCN; Young and Calhoun 2005; Yu and Young 2000), but their performance degrades at higher levels of the auditory system where the neurons become more nonlinear (Eggermont et al. 1983a; Escabi and Schreiner 2002; Machens et al. 2004; Nagel and Doupe 2006; Nelken et al. 1997; Sen et al. 2001; Theunissen et al. 2000; Versnel and Shamma 1998; Yeshurun et al. 1989). Here we investigate the nature of the nonlinearity for principal neurons of the dorsal cochlear nucleus (DCN) by considering a type of stimulus for which the responses can be accurately linearized by varying the spectral contrast.

In this analysis, the STRF is simplified to a function of frequency by averaging over the time dimension (Young et al. 2005). The resulting spectral receptive field is a model of the relationship between the spectral shape, or frequency content, of a stationary stimulus and the average discharge rate of the neuron, without regard to temporal aspects of either the stimulus or the response. The first-order spectral receptive field assumes that the discharge rate is given by a weighted summation of the stimulus spectrum. Simplification to the frequency dimension allows the model to be extended to second order, meaning that effects of the interaction of pairs of frequencies are also included (subsequently called a “quadratic” model). When a quadratic model is used to predict responses to filtered noise stimuli based on their spectral shapes, the second-order terms improve the fit (Young and Calhoun 2005; Yu 2003; Yu and Young 2000). However, for nonlinear neurons in the DCN even the second-order model is unable to fully account for rate variations.

A relatively unexplored aspect of auditory stimuli is their spectral contrast, meaning the range of variation of the stimulus spectrum through time (Kvale and Schreiner 2004; Nagel and Doupe 2006) or across frequency (Barbour and Wang 2003a,b; Calhoun and Schreiner 1998; Escabi et al. 2003); contrast can be measured, for example, by the SD of the power spectrum or the stimulus envelope. Changes in spectral contrast can produce changes in receptive fields of auditory neurons or induce adaptation and changes in neural “gain.” For this paper, it is most important that the ability of STRFs or spectral models to accurately predict the responses of neurons is expected to vary with the spectral contrast. Because models are approximations of the neuron’s input–output function, at the costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Address for reprint requests and other correspondence: E. D. Young, Department of Biomedical Engineering, Johns Hopkins University, 720 Rutland Ave., Baltimore, MD 21205 (E-mail: eyoung@jhu.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

www.jn.org 0022-3077/07 $8.00 Copyright © 2007 The American Physiological Society
lower spectral contrast a smaller range of that function needs to be approximated and the model should do better. Nevertheless, natural sounds have spectral contrasts that are more like the 12-dB value that gives poor performance from models (Attias and Schreiner 1997; Escabi et al. 2003; Singh and Theunissen 2003), so it is important to understand the extent to which receptive fields can be linearized using small contrast and the extent to which linearized receptive fields are representative of those obtained at natural levels of contrast.

Here, we consider the effects of spectral contrast on the ability of the quadratic model to fit the responses of so-called type IV neurons in DCN. As expected, the models work better at lower spectral contrast. The model weights also typically increase in magnitude at lower spectral contrast. Similar effects on weight size are observed in AN fibers and may reflect cochlear compression and transducer nonlinearity. Importantly, the general shapes of the models are similar at low and high spectral contrast, suggesting that even when models do not accurately predict responses, they may capture some meaningful properties of the neuron.

**METHODS**

**Surgical procedures**

Experiments were conducted on a total of 13 adult cats (3–4 kg) with infection-free ears and clear tympanic membranes. Animal-use protocols were approved by the Johns Hopkins Animal Care and Use Committee. For DCN recording, 11 cats were anesthetized with xylazine [2 mg, administered intramuscularly (im)] plus ketamine (40 mg/kg im, supplemental dose: 15 mg/kg im). Atropine (0.1 mg im) was given to control mucous secretion. A tracheal tube was inserted. Cats were decerebrated by aspirating through the brain stem between the superior colliculus and thalamus, after which anesthesia was discontinued. Core body temperature was maintained at about 38°C using a regulated heating blanket and lactated Ringer solution was given intravenously (iv) to maintain fluid volume.

The DCN was exposed by drilling a hole posteriorly through the bone lateral to the foramen magnum, reflecting the dura, and gently parting the choroid plexus. Recording electrodes were advanced into the DCN under visual control. Single neurons were isolated and recorded extracellularly using platinum–iridium microelectrodes. Action potentials were detected with a Schmitt trigger and spike times recorded with a precision of 10 μs.

In two additional experiments, cats were prepared for recording in the AN, as described previously (Young and Calhoun 2005). These animals were treated as described earlier, except that they were kept anesthetized (sodium pentobarbital, 5–15 mg/h iv, to effect) and were not decerebrated. The cerebellum was retracted to expose the AN and fibers were recorded with pipette electrodes.

**Experimental protocol and numbers of neurons**

Recordings were made in a sound-attenuating chamber. Acoustic stimuli were delivered to the ipsilateral ear by an electrostatic speaker coupled to a hollow ear bar. The speaker was calibrated in situ using a probe tube placed approximately 2 mm from the eardrum. The calibration is essentially flat with fluctuations of <10 dB from 0.5 to 30 kHz.

Once a DCN neuron was isolated, it was characterized using a combination of tones and broadband noise (BBN). Rate versus level functions were collected for best-frequency (BF) tones and BBN by presenting 200-ms-stimulus bursts (10-ms rise/fall times) once per second over an 80- to 100-dB range of sound levels in 1-dB steps. Generally only type IV neurons were studied; these are the most common principal cell type in the DCN in decerebrate cats (Young 1980). Type IV neurons were classified as having moderate spontaneous rates and BF-tone rate-level functions with excitation at low sound levels and inhibition at high sound levels (Shofner and Young 1985). Only neurons located along the electrode track before a BF gradient shift, which indicates a transition from DCN to VCN, were considered to be DCN neurons. Two type IV neurons are included that may have been located in VCN. Their properties are otherwise typical of DCN neurons. Two type II neurons were also studied briefly; these had spontaneous rates near zero and showed strongly excitatory but nonmonotonic responses to BF tones, but very little response to noise (<30% of maximum BF tone rate); they had higher thresholds than those of type IV neurons, and were found by searching with BF tones 20–30 dB above those of type IV thresholds.

In summary, a total of 14 neurons in the CN were studied: 12 type IV neurons and 2 type II neurons. Of the 12 type IV neurons, 10 were classified as DCN type IVs and the remaining two were classified as ambiguous in location (either DCN or VCN). Some type IV neurons were recorded at multiple sound levels, so that the total data set consists of 17 cases, including some neurons at different sound levels. Data from the two neurons with ambiguous locations are included in all the summary diagrams because their behavior was essentially the same as that of clear DCN neurons. Data from the type II neurons are not shown, although they are briefly described. The yield of neurons per experiment is small because of the long times needed to collect sufficient repetitions for each neuron (several hours; see the last paragraph of the section RSS stimuli) and because these experiments were used to collect data for other protocols as well (published in Reiss 2005; Reiss and Young 2005; Young et al. 2005). Moreover, additional neurons were sampled but not included because of fragmentary results. Those data are consistent with the data reported here in every way.

In the AN experiments, fibers were characterized as to BF, Q10, and spontaneous discharge rate. Data are included from 14 fibers in which complete data sets were obtained. Again, the yield is low because these experiments were used to obtain control data for this and another experiment.

The acoustic stimuli described in the next section were presented at a rate of one stimulus per second, with the first stimulus presented an octave, and all eight tones within a bin have the same amplitude; the sound level \( S(f) \) in each frequency bin is the sum of the powers of each tone. The acoustic stimuli consisted of 17 cases, including some neurons at different sound levels. Each stimulus was presented over a range of sound levels, spaced at 10 dB and beginning 5–20 dB above threshold. Response rate was computed as the number of spikes over the stimulus duration divided by the duration.

**RSS stimuli**

The random spectral shape (RSS) stimuli used here are similar to those used before (Young and Calhoun 2005; Yu and Young 2000). Each stimulus consists of a sum of tones spaced logarithmically at 1/64th octave. The tones are grouped into frequency bins of 1/8th octave, and all eight tones within a bin have the same amplitude; the sound level \( S(f) \) in each frequency bin is the sum of the powers of these tones. The starting phases of the tones were randomized to avoid a click at stimulus onset; the same phase set was used for all stimuli. Linear 10-ms onset and offset ramps were added to the time-domain signal. These stimuli were not corrected for spectral irregularities in the speaker calibration.

The RSS stimuli had a bandwidth of 6 octaves, centered on 6 kHz. Each RSS set consisted of 400 stimuli; in 392 of those, the amplitudes (in dB) of the bins \( S(f) \) were selected pseudorandomly from an approximately Gaussian distribution with mean 0 dB. \( S(f) \) is the dB level, relative to a reference sound level, of the sound in the bin centered on frequency \( f \). The remaining eight stimuli had \( S(f) = 0 \) dB for all \( f \), i.e., the reference sound level in each bin; these all-0-dB stimuli were uniformly scattered through the other stimuli. Stimuli were organized into successive plus-minus pairs, so that the dB levels
of the first stimulus of each pair \( S_i(f) \) were inverted in the second stimulus \( S_{-i}(f) = -S_i(f) \). These plus-minus pairs were used to separate the estimation of even- and odd-order terms, described in the next section. The eight all-0-dB stimuli were used to estimate the reference rate \( R_0 \). Note that the all-0-dB stimuli are not “flat-spectrum,” in the usual sense. Because of the logarithmic spacing of tones, this spectrum actually has a 1/\( f \) shape.

We use the SD of the distribution of spectral amplitudes (in dB) as the measure of spectral contrast. Previous studies from this lab used a constant spectral contrast of 12 dB (Young and Calhoun 2005; Yu and Young 2000). In this study, spectral contrasts were varied across sets to be 12, 6, or 3 dB. As shown in Fig. 1, the underlying spectral shape is the same across contrast, but the dB levels in each bin are scaled to achieve the desired SD. Except where noted, the stimulus sets with different spectral contrasts were presented separately.

To compare fairly the model performance in response prediction across spectral contrasts, it was necessary to equalize the effects of the random noise in the discharge-rate estimation. When the spectral contrast is low, rate changes are small and more repetitions are needed to achieve a particular signal-to-noise ratio in estimating the model parameters and test response rates. It can be shown that if the response is determined by the first-order weights only and the random fluctuation in discharge rates is Gaussian with variance proportional to the spike count, consistent with Poisson spike counts, then the SDs of the weight estimates are inversely proportional to the square of the stimulus SD. Therefore to achieve comparable errors in weight estimation, the 6-dB stimulus set must be repeated four times for every repetition of the 12-dB set and the 3-dB set must be repeated 16 times. This is a lower bound on the required number of repetitions because nonlinear weights further increase the difference in expected SD between contrasts. Thus data were included only if multiple repetitions of data acquisition for the 3- and 6-dB sets were obtained; the lower bounds of required repetitions for all spectral contrasts were achieved whenever contact with the neuron was maintained long enough (12/17 cases). The remaining 5/17 cases were included because the number of repetitions was close to this lower bound.

For the AN fibers, data were obtained only at 3- and 12-dB spectral contrast with only one repetition of each stimulus because of limited recording time. As a result, the SD of rate estimates is larger and the quality of prediction at 3-dB contrast is lower because of the added variance in the rate estimates. The RSS sets used here differed slightly from those used in DCN, as described in Bandyopadhyay, Reiss, and Young (unpublished observations). The main difference is that the 0-dB stimuli were placed at the end instead of interleaved within the main set, and typically only half the full data set (200 stimuli) was recorded for AN fibers. Therefore for the AN data, \( R_0 \) was estimated as part of the parameters instead of from the all-0-dB stimuli as in the DCN.

**Quadratic weighting function model**

The average discharge rate \( r \) is modeled using a quadratic gain function as follows

\[
r = R_0 + \sum_{j} w_j S_j(f) + \sum_{j} \sum_{k} m_{jk} S_j(f) S_k(f) \tag{1a}
\]

or the equivalent

\[
r = R_0 + w^s s + s^2 M s \tag{1b}
\]

where \( R_0 \) is the reference discharge rate, i.e., the rate in response to an all-0-dB stimulus; \( s \) is a vector containing the dB values of the stimulus at different frequencies, i.e., \( s = [S_1(f), S_2(f), \ldots, S_N(f)]^T \); \( w \) is a vector of first-order weights \( w_j \), where each weight is the “gain” of the neuron in spikes/(s·dB) for sound energy in the corresponding frequency bin; and \( M \) is a matrix of second-order weights \( m_{jk} \), each of which is the gain of the neuron in spikes/(s·dB\(^2\)) for joint sound energy in the \( j \)th and \( k \)th bins.

The motivation for this model has been discussed previously (Young and Calhoun 2005). One important aspect of the model is the use of a logarithmic (i.e., decibel) measure of stimulus intensity. This is different from most STRF models; however, recent STRF analyses have shown improved performance when logarithmic compression (i.e., a decibel scale) is used in measuring the stimulus intensity (Escabi et al. 2003; Gill et al. 2006).

To estimate the parameters of the model, rates are measured in response to the RSS set described in the previous section, giving 392 equations like Eq. 1, one for each stimulus in the set. The reference rate \( R_0 \) is the average rate of the eight all-0-dB stimuli. The weights are estimated by minimizing the chi-square error of the model as

\[
\chi^2(w;M) = \sum_j \left[ \frac{r_j - \hat{r}_j(s,w;M)}{\sigma_j} \right]^2 \tag{2}
\]

where \( r_j \) represents the rates measured in the experiment and \( \hat{r}_j(s,w;M) \) is the rate predicted by the model for stimulus \( s \) with the weights \( w \) and \( M \). The SDs of the rates \( \sigma_j \) could not always be estimated from the data because the stimuli were not always repeated, so the SDs were computed assuming Poisson spike counts, i.e., \( \sigma_j^2 = r_j/T \), where \( T \) is the duration of the stimulus. The Poisson assumption was tested using data from 15 DCN type IV neurons by plotting spike count variance versus mean (for responses to RSS stimuli). For most cases, the neural data are within the statistical limits (99% confidence) of Poisson spike counts, given the same rates and durations; the major exception is at high discharge rates, where the variance is smaller than expected, probably because of neural refractoriness. To avoid dividing by zero, \( \sigma_j^2 \) was set to 1 whenever it was <1. The estimation was done by the method of normal equations (Press et al. 1992) and the SDs of the weight estimates were computed from bootstrap repetitions of the

---

**FIG. 1.** An example random spectral shape (RSS) spectrum at 3 spectral contrasts—12, 6, and 3 dB—from top to bottom. Each frequency bin is 1/8th octave wide with 8 frequency components (tones) in the bin. Stimulation was 6 octaves wide centered at 6 kHz.
calculation. The bootstrap calculation probably overestimates the SD (Young and Calhoun 2005).

The design of the plus-minus stimulus set allowed the weight estimation to be done separately for the odd- and even-order weights. Suppose \( r^+ \) and \( r^- \) are the rates in response to a plus-minus stimulus pair \( s^+ \) and \( s^- \), where \( s^+ = -s^- \); then from Eq. 1

\[
\frac{r^+ - r^-}{2} = R_0 + s^T M s \quad \text{and} \quad \frac{r^+ + r^-}{2} = s^T W
\]

With this method, the estimation of \( M \) is done from \((r^+ + r^-)/2\) and \( W \) is estimated from \((r^+ - r^-)/2\). \( R_0 \) is estimated separately from the responses to all-0-dB stimuli. This method is useful because the estimates of different orders of weights are not orthogonal, so if the neuron were to contain a significant third- (or higher) order response, it would appear as an error in estimation of lower-order weights. Estimating even- and odd-order weights separately reduces this error by keeping the error from unestimated odd-order components from affecting even-order estimates and vice versa.

To maximize the ratio of data to parameters, weights were computed for a limited number of frequency bins around the BF. This frequency range was chosen by estimating first-order weights over the full frequency range and selecting a continuous range of frequency bins with significant weights, i.e., weights >1 (bootstrap) SD from zero. If two significant bins were separated by a nonsignificant bin, the range was made continuous by including the nonsignificant bin. First- and second-order weights were then estimated over the significant range, subject to the constraint that the number of equations was \( \geq 2.5 \) times the number of weights. Weights were excluded at the upper or lower frequency limits to achieve this criterion.

Testing validity and generality of the model

To test the quality of the fit, the model was estimated from 75% of the data points and then tested by using it to predict the remaining 25% of the data points. Confidence intervals were calculated using 1,000 bootstraps (Efron and Tibshirani 1993), each time estimating the model from a randomly chosen 75% of the data and predicting responses to the remaining 25% of stimuli. The measure of fitting accuracy was the fraction of variance, defined as

\[
fv = 1 - \frac{\sum (r_i - \hat{r}_i)^2}{\sum (r_i - \bar{r})^2}
\]

The symbols are as defined in Eq. 2 except that \( \bar{r} \) is the mean rate. \( fv \) has a maximum value of 1, when the model fits the data perfectly, and decreases as the error increases. It is zero when the mean rate fits as well as the model, and can go negative for poor models. Here, \( fv \) was not limited at zero. The estimate of \( fv \) is decreased by random noise in the estimates \( r_i \). Variance corrections are sometimes used in an attempt to control this effect. However, those corrections are somewhat unstable. Given the purpose of this paper, to compare models at different stimulus contrasts, we did not use the variance correction. The use of multiple repetitions of the stimulus at small contrast, described earlier, should serve to counteract this effect. However, because multiple repetition of stimuli was not done in the AN experiments, the estimates of \( fv \) for 3-dB contrast are strongly affected by noise in the rates.

The Pearson product-moment correlation coefficient \( (r) \), computed between \( r_i \) and \( \hat{r}_i \), is often used as a measure of goodness of fit. For our data, \( r^2 = fv \); however, \( r \) and \( fv \) are differentially sensitive to errors in the mean rate, with \( fv \) being more sensitive. Because of this and other effects, the \( fv \) is a more conservative measure of goodness of fit; that is, \( fv \) and \( r^2 \) computed from the same data and fits reliably give \( r^2 \) values that are slightly larger than \( fv \), for both the AN and DCN data shown here. Because of this difference, we have chosen to use \( fv \) for comparison with other work, a good rule of thumb is that the correlation coefficient is equal to or larger than \( fv^{1/2} \).

Eigenvectors of second-order weights

In dealing with the second-order weights, it is useful to express \( M \) in terms of its eigenvectors as follows. \( M \) is a real, symmetric \( N \times N \) matrix and can be decomposed into its \( N \) eigenvectors \( x_i \) with their corresponding eigenvalues \( \lambda_i \), as follows

\[
M = \sum_{i=1}^{N} \lambda_i x_i x_i^T
\]

The eigenvectors are frequency-specific weight vectors like \( w \), in that the second-order response to a stimulus \( s \) can be written as

\[
s^T M s = \sum_{i=1}^{N} \lambda_i (s^T x_i)^2
\]

The \( i \)th eigenvector weights the stimulus spectrum and contributes an excitatory or inhibitory effect, according to the sign of \( \lambda_i \).

RESULTS

Quadratic spectral processing model

The quadratic model (Eq. 1) was fit to the responses of 12 type IV neurons for various values of spectral SD (contrast). In each case, two changes in the model were observed when the spectral contrast was changed: first, the weights were typically larger at smaller spectral contrasts; second, the model performed better in prediction tests at lower spectral contrast. Note that “prediction” means fitting the model to 75% of the data points and then testing it by predicting the responses to the remaining 25% of data.

An example of the weights at three spectral contrasts is shown for one DCN neuron in Fig. 2. The first- and second-order weights are shown in Fig. 2A and B. Clearly, the weights are larger in amplitude at lower spectral contrast, even though the weights have the same general shape at all contrasts. For example, the first-order weights (Fig. 2A) are predominantly inhibitory with the inhibition centered just below BF at all three contrasts. The second-order weights (Fig. 2B) are typical of those in type IV neurons (Yu 2003) in that the weights are negative on the diagonal of the weight matrix [e.g., for same-frequency terms like \( m_{jj} S(f_j)^2 \)] and positive off-diagonal.

The effects of the second-order weights are best visualized on the basis of the eigenvalues and eigenvectors of \( M \), plotted in Fig. 2C. The eigenvectors corresponding to the largest positive (left) and negative (right) eigenvalues are shown, plotted as eigenvector multiplied by eigenvalue. These vectors define specific spectral shapes in the stimulus that are excitatory or inhibitory depending on the sign of the corresponding eigenvalue. Note that the eigenvector-stimulus operation is squared in Eq. 6, so the effects of a particular eigenvector are independent of the absolute sign of the stimulus, i.e., they are the same for \( s \) and \(-s\). In this case (Fig. 2C), the positive eigenvalues contribute little to the response, and the largest negative eigenvalue contributes an inhibitory effect for a complex spectral shape.

The inset to the plots in Fig. 2C gives the eigenvalues at the three contrasts. These increase as the spectral con-
Effects of decreased contrast in the population

The prediction performance for 17 cases studied in 12 type IV neurons at a range of sound levels is summarized in Fig. 4A. Prediction improved (larger $f_v$) with decreased contrast for all cases at all sound levels.

Weight amplitudes also increased consistently at lower contrast. When a sufficient number of repetitions was collected (see METHODS), clear weight increases were observed in 11/11 cases at low-intermediate sound levels ($-10$ to 30 dB), and in 5/6 cases at high sound levels (>30 dB SPL). The weight amplitudes were quantified for first-order weights by calculating the norms of the weight vectors (Fig. 4B) and for second-order weights by using the largest (positive) and smallest (largest negative) eigenvalues of $M$ (Fig. 4D). The other eigenvalues are not shown because they generally showed the same trends. Overall the median first-order norms were 1.7 and 5.1 sp/(s·dB) at 12- and 3-dB contrast, significantly different at

A comparison of weight magnitude at 3-dB contrast decreases, reflecting the increase in the weight values in Fig. 2B.

An example of the quality of the fits is shown in Fig. 2D, for both the fitted data (the 75% of data to which the model was fitted, blue) and the prediction data (the remaining 25% of data, red). The model does better with both the fitted and predicted rates at smaller contrast; for the predictions shown in Fig. 2D, the $f_v$ values are 0.33 (12 dB), 0.87 (6 dB), and 0.92 (3 dB).

The weights and the model performance for the same neuron at 20-dB lower stimulus level are shown in Fig. 3. As is typical, the shapes of the weight functions change with overall sound level (Yu 2003). Note that overall sound level, called the

---

FIG. 2. Quadratic model for a dorsal cochlear nucleus (DCN) type IV neuron with best frequency (BF) = 13.4 kHz, studied at a reference level of 35 dB SPL/component, i.e., each tone is at 35 dB SPL. $A$: 1st-order weights at 3 spectral contrasts, identified in the legend. Weights are plotted as a function of the center frequency of the RSS bin, expressed as octaves relative to the BF of the neuron. $B$: corresponding 2nd-order weight matrices, plotted on 2 frequency axes. Color scale for weight amplitude is the same in the 3 plots and is identified at right. $C$: eigenvectors ($x_i$ in Eqs. 5 and 6) corresponding to the largest positive (left) and negative (right) eigenvalues of the 2nd-order weight matrix. Eigenvectors are plotted as eigenvalue multiplied by eigenvector, and the eigenvalues are given on the plots (these have units spikes/(s·dB²)). Plots are identified by the same line styles as in $A$. $D$: plots of the rates from the data (abscissa) vs. the rates predicted by the quadratic model (ordinate) to show the quality of the fit; the lines show where the model predictions equal the data. Results are shown for all 3 contrasts (identified in the legends) and for both the fit (the 75% of the data used to compute the model parameters) and the prediction (the remaining 25% of the data), identified by symbol color. Note that these results are not (necessarily) a bootstrap trial, but are taken from data divided as just described.

FIG. 3. Same results as in Fig. 2 for a lower reference sound level (15 dB SPL/component) in the same neuron. In $C$, the eigenvector for 12-dB spectral contrast is small and hard to discern.
As in the examples in Figs. 2 and 3, the weights changed primarily in amplitude, and not in shape, as spectral contrast changed. The similarities of weight shape for 12- and 3-dB contrasts were quantified by the cross-correlation or similarity index (SI)

$$SI = \frac{\langle x_{12}, x_i \rangle}{\|x_{12}\| \cdot \|x_i\|} = \frac{\sum_{i} x_{12}[i]x_i[i]}{\sqrt{\sum_{i} (x_{12}[i])^2 \cdot \sum_{i} (x_i[i])^2}}$$

(7)

where $x_{12}$ and $x_i$ are the 12- and 3-dB weight vectors (or eigenvectors) for a neuron. The SI measures the cosine of the angle between the vectors; SI = 1 for vectors that differ only in length. Figure 4C shows that the majority of neurons had first-order weight vectors at the 12- and 3-dB spectral contrast with SI values ~1. A similar result is seen for the largest-eigenvalue eigenvectors of the second-order weight matrices (Fig. 4E), except that the similarities are smaller and the variability in the measures is larger.

**Compression of the range of discharge rates**

The decrease in weight size as spectral contrast increases should result in a compression of the range of discharge rates produced by stimuli at higher spectral contrast. The examples in Figs. 2D and 3D show this effect, in that the range of discharge rates produced by the RSS set is smaller at larger contrast than would be predicted by a linear model with fixed gain. For example, going from the 3- to the 12-dB contrast increases the amplitudes of stimulus bins by a factor of 4, but significantly increases the range of rates by a factor of <4. The range of rates produced by the RSS sets was measured as the interdecile range, the range between the 10 and 90% points in the populations of rates along the abscissae of Figs. 2D and 3D. The interdecile ranges are plotted in Fig. 5 as the range for the

![Graph showing population summaries for the quadratic model for 17 cases studied in 12 type IV neurons. Some neurons were studied at more than one sound level and are represented by more than one point in each plot. Reference sound level of the stimuli is indicated by the color and symbol style, according to the legend at top right (dB SPL per component). In all panels, the points show the mean of 1,000 bootstrap trials and the vertical lines show the SD. A: fv (Eq. 4) for prediction data (ordinate) vs. spectral contrast (abscissa). Two neurons had fv values <0 at 12-dB contrast and those points are off-scale. B: norm (length) of the 1st-order weight vector is plotted vs. spectral contrast for the same data. C: similarity index (Eq. 7) between the 1st-order weight vectors at 3- and 12-dB contrast are shown for each case, indexed along the abscissa in the order in which the neurons were studied. D: largest positive and negative eigenvalues of the 2nd-order weight matrix are plotted vs. spectral contrast. E: similarity between largest-eigenvalue eigenvectors at 12- and 3-dB contrast for each case. Results for the largest positive eigenvalue are in the top plot and for the largest negative eigenvalue in the bottom plot.

$P \approx 10^{-4}$. The median ratio of the 3- and 12-dB first-order norms was 2.7. Similar results were obtained for the eigenvalues, with median ratios of 4.0 and 3.7 for positive and negative eigenvalues, respectively.

As in the examples in Figs. 2 and 3, the weights changed primarily in amplitude, and not in shape, as spectral contrast changed. The similarities of weight shape for 12- and 3-dB contrasts were quantified by the cross-correlation or similarity index (SI)
The static nonlinearities were applied to the models at 12-dB contrast. As seen in Fig. 6, they did not improve the prediction performance of the model. Figure 6A compares the $f_v$ of the prediction performance for the quadratic model without the static nonlinearity (absicssa) and the $f_v$ including the static nonlinearity (ordinate). With all three nonlinearities (see legend) the performance is worse with the nonlinearity. Of course, the fit of the model to the data was slightly better with the static nonlinearities (Reiss 2005), even though the prediction performance did not improve. This result probably derives from overfitting of the data by the static nonlinearities.

Consistent with this negative result, adding the static nonlinearity had only small effects on the weight amplitudes at 12-dB contrast (Fig. 6, B and C); the lengths (norms and eigenvalues) of weight vectors were only slightly larger with the static nonlinearity.

**Effects of spectral contrast in the auditory nerve**

It is useful to compare the effects of stimulus contrast on DCN responses with the effects seen in the AN. Sufficient data were obtained from 14 AN fibers with BFs from 2.05 to 9.07 kHz and a range of spontaneous rates (0.5–95/s). Data were obtained at 3- and 12-dB contrast only; as noted in METHODS, repeat presentations of the 3-dB stimuli were not done because of limited recording time. Generally the results from the AN fibers are qualitatively similar to those from DCN neurons.

Data from an example AN fiber are shown in Fig. 7 and the population summary of AN fibers is shown in Fig. 8. As in the DCN, the first- and second-order weights of AN fibers were larger for the 3-dB stimuli than those for the 12-dB stimuli (Fig. 7, A–C); this result also held for the population of AN fibers (Fig. 8, A and B). Quantitatively, the median ratio of the norms of the first-order weights for 3-dB relative to 12-dB stimuli was 1.6 (significantly different from 1 at $P < 0.001$, two-sided sign test). This compares to 2.7 in DCN neurons (significantly different at $P < 0.002$, rank-sum test). Second-order weights were also larger for 3-dB stimuli in the AN fibers; the median ratio of the largest eigenvalues, 3 dB over 12 dB, was 5.0 for positive and 13.0 for negative eigenvalues (both significantly different from 1 at $P < 10^{-5}$). These numbers are noticeably larger than the ratios in the DCN (4.0 for positive eigenvalues, 3.7 for negative eigenvalues). This result reflects mainly the very small size of second-order eigenvalues for the 12-dB contrast stimuli in the AN data.

The shapes of the AN weight functions did not change much between 3- and 12-dB stimuli, as indicated by the similarity indices (Fig. 8, C and D). The similarity indices were generally smaller for the second-order weights in the AN compared with the DCN data, again reflecting the small size of the second-order weights at 12-dB contrast.

The behavior of the rate range was similar to that for DCN neurons (Fig. 5, + symbols), with the 3-dB range somewhat larger than expected if there were no gain change. However, the changes in the weights from 3- to 12-dB contrasts suggest that the adjustment of rate ranges depends more on second-order terms in the AN, compared with the DCN.

The most noticeable change from DCN neurons to AN fibers is the increase in prediction performance at 12-dB contrast in 12-dB contrast stimuli (abscissa) versus the ranges for the 3- and 6-dB contrasts (ordinate). The dashed line shows the expected location of the data points for 6-dB contrast (filled circles), based on a linear model with constant weights; in this condition, the rate ranges for 6-dB contrast should be half the values for 12-dB contrast. The dotted line shows the ratio of 1/4 expected for the 3-dB contrasts. The points for both the 3- and 6-dB contrasts are all above the lines, except for one point, showing that rate ranges are compressed at higher contrasts. The pluses show AN data, subsequently discussed.

**Saturation and threshold nonlinearities**

Conceivably, these changes in weight size could occur if the range of rates produced by the neuron is constrained by saturation at high rates and by a threshold at zero rate, or the spontaneous rate. This seems unlikely in our data because the rate plots in Figs. 2D and 3D are typical of DCN type IV neurons in that they do not show obvious rate constraint effects, except at zero rate.

To evaluate the importance of a static nonlinearity, three static nonlinearities were incorporated into the model: a simple one-parameter rectifier with a threshold; a two-parameter rectifier with a threshold and saturation; and a sigmoidal rate function given by $g(x) = a/[1 + \exp(-bx + c)]$. The static nonlinearities were placed after the quadratic model as in previous studies (e.g., Chander and Chichilnisky 2001; Nagel and Doupe 2006). The parameters of each nonlinearity were estimated concurrently with the weights, by combining them into a single-parameter vector (containing the $w_j$, the $m_{jk}$, and the parameters of the nonlinearity) and then using a gradient descent algorithm (Nelder–Mead simplex direct search using the Matlab function fminsearch) to minimize the chi-square error function (Eq. 2). The original weighting function parameters were used as seed parameters to improve convergence.

FIG. 6. Results of adding 3 different static nonlinearities to the model are shown. A: prediction performance ($f_v$) of the model with (ordinate) and without (abscissa) a static nonlinearity is shown. Symbols identify the type of nonlinearity (identified in the legend at right) and the lines show bootstrap SDs of the estimates of $f_v$. B: length (norm) of 1st-order weight vectors with and without the static nonlinearity. Symbols and lines are the same as in A. C: maximum and minimum eigenvalues of 2nd-order weight matrices estimated with or without a static nonlinearity.
the latter. This change is clear in the example of Fig. 7D, where the \( f_v \) values are 0.33 (3 dB) and 0.69 (12 dB). The median \( f_v \) for the AN population at 12-dB contrast was 0.59, similar to the value obtained previously in AN fibers (0.59; Young and Calhoun 2005) and substantially larger than the median value for type IV neurons (~0.4, significantly different at \( P = 0.08 \), rank-sum test). Prediction performance is not shown for the AN population because only one repeat of the 3-dB stimuli was obtained. As a result, \( f_v \) values for the 3-dB stimuli are decreased by the noise in the estimation of relatively small rate changes from only one repetition of the stimulus. In fact, \( f_v \) values were often smaller for the 3-dB stimuli than those for the 12-dB stimuli (median value 0.30).

The nature of the contrast effect in the AN was investigated further by computing responses to 3- and 12-dB contrast RSS stimuli using a cochlear model (Bruce et al. 2003). This model contains two nonlinearities (Carney 1993). First, there is a level-dependent filter that decreases its gain and increases its bandwidth as stimulus amplitude increases; it is meant to model fast cochlear compression due to outer hair cell function (Robles and Ruggero 2001). Second, there is a static nonlinearity representing the inner hair cell input–output function. This function saturates with both hyperpolarization and depolarization. For the model data, the first-order weights and the positive second-order eigenvalues of the quadratic model were larger at lower contrast, as expected (Reiss 2005); this behavior was not observed for near-threshold stimuli where responses to the 3-dB stimuli were very weak. Prediction performance was very good for both contrasts (with repeats; \( f_v > 0.8 \)), and it was slightly better for 3-dB contrast.

Reducing either nonlinearity in the model reduced the differences between the quadratic models of responses to the 3- and 12-dB stimuli. Thus the model results suggest that contrast gain changes seen in the AN are due to both cochlear compression and the static input–output nonlinearities in the transduction path.

Consistent with the suggestion that contrast gain changes in CN may be inherited in part from the cochlea, two DCN type II neurons (an inhibitory interneuron in DCN; Young and Davis 2002) were also studied and both showed the same effects of stimulus contrast as did principal cells. Specifically, they had lower gain at higher contrast.

**DISCUSSION**

As RSS spectral contrast was decreased from 12 to 3 dB, the quality of the representation of type IV responses by the quadratic model improved, as shown by better prediction.
performance (Fig. 4A). This result was expected because non-linear systems can usually be modeled with linear approximations for sufficiently small deviations about an operating point. We also found that both the first-order weights and the second-order eigenvalues increased in magnitude, but that weight functions maintained similar shape, as spectral contrast was decreased (Fig. 4, B–E). The changes with contrast seen in the DCN are also seen qualitatively in the AN, except for the decrease in prediction performance at high contrast (Fig. 8). In considering the mechanisms of these effects, it is necessary to consider which of the results in DCN are a result of processes in the cochlea and which are emergent properties of the cochlear nucleus.

One possible interpretation of the change of weight magnitudes with contrast is that this effect represents a form of contrast gain control, similar to that seen in visual neurons (e.g., Chander and Chichilnisky 2001; Enroth Cugell and Lennie 1975; Shapley and Victor 1978); that is, the gain of auditory neurons for stimulus modulation is adjusted for the amplitude range of the modulation, as measured by contrast or variance of the envelope (Bonin et al. 2006; Zaghoul et al. 2005). An ideal contrast gain control would maintain the range of discharge rates fixed as the variance increases or decreases. Figure 5 shows that contrast gain control in DCN neurons and AN fibers does not conform to ideal gain control. Just as all the points (barring one) for 3 and 6 dB in Fig. 5 are above the dotted and dashed lines, respectively, they are also all below the solid line (equality), which corresponds to perfect gain control.

**Other studies of stimulus contrast in auditory neurons**

The effects of spectral contrast have been studied in cat auditory cortex using auditory gratings or ripple stimuli (Calhoun and Schreiner 1998). In these neurons, varying the ripple modulation depth (analogous to spectral contrast) often caused nonlinear changes in the ripple modulation transfer function, estimated from measurements of rate responses to several ripple densities. These results suggest that spectral contrast influences the shapes of spectral receptive fields, unlike the effects shown here, which generally show good similarity between first-order weight-function shapes at different contrasts. However, modulation depth changes both spectral contrast and overall sound level so the receptive-field changes could have resulted from either change in the stimulus.

Spectral contrast has also been studied in marmoset auditory cortical neurons using similar RSS stimuli (Barbour and Wang 2003a,b). In many neurons the results were similar to those described here: a decrease in spectral contrast resulted in an increase in first-order weights and little change in tuning across frequency; prediction performance was low. However, unlike the results here, many neurons showed contrast preference, with both low- and high-contrast preferring neurons. A similar result was obtained by Escabi and colleagues (2003) in the inferior colliculus, although they defined contrast differently. Contrast preference is not a feature of the neurons studied here and seems to be a property of the auditory system that develops at a level above the cochlear nucleus.

Possibly related effects of stimulus contrast in the temporal domain have been studied in the inferior colliculus for sinusoidal carriers modulated by Gaussian noise or m-sequences (Kvale and Schreiner 2004). Adaptive effects of a sudden change in contrast were observed in which an increase in contrast resulted in a decrease in the gain of the neuron for amplitude modulation (AM), an effect that is qualitatively similar to the changes in weight amplitudes observed here. Similar results were obtained in neurons in avian field L for stimuli consisting of AM of a noise carrier by band-pass noise (Nagel and Doupe 2006). Sudden changes in the contrast (variance) of the envelope led to changes in the gain of the neuron for the modulation signal, such that a decrease in the gain was observed for stimuli with larger contrast. The effects in both cases were adaptive, in that the changes in gain occurred over a few hundreds of milliseconds after the change in contrast in response to continued stimulation. In this aspect the temporal contrast responses differ from those studied here (see next section). Also, as part of the adaptation process, the temporal filtering properties of the modulation response showed consistent but small changes in the colliculus but not in field L.

**Causes of changes in weighting function gain with spectral contrast in DCN**

The data in Figs. 7 and 8 show that all of the effects observed in type IV neurons with changes in stimulus contrast are also seen in AN fibers, except for the poor prediction performance at 12-dB contrast. The fact that a nonlinear cochlear model shows the same effects as AN fibers, discussed in RESULTS, suggests that the decrease in AN gain with increased contrast is a result of fast cochlear compression, represented in the model by the saturation of the inner-hair cell model and the level-dependent gain of the outer-hair cell model.

However, it is unlikely that the cochlear effect fully accounts for the changes in gain seen in type IV neurons. First, the AN data fail to quantitatively account for the effects seen in type IV neurons. Second, there is a substantial increase in the degree of nonlinearity in going from AN to type IV neurons. Considering the first-order weights, the increase in weight magnitude in going from 12- to 3-dB contrast is larger in type IV neurons than that in AN fibers, suggesting a change in weight magnitudes within the cochlear nucleus. Considering the second-order weights, the ratios of eigenvalues of the second-order weight matrix are the reverse, larger in the AN than in the type IV neurons. The large second-order ratios in the AN seem to be a result mainly of very small second-order weights in AN fibers at 12-dB contrast, a reflection of the linearity of AN fiber responses at large contrast. This suggests that the relatively larger second-order weights in type IV neurons at large contrast are primarily an effect of increased nonlinearity in the cochlear nucleus.

The gain change with contrast studied here is unlikely to be a slow adaptive process, like those discussed earlier for the inferior colliculus and field L neurons, in which the neurons adjust their gains based on some properties of the preceding stimuli. First, such a mechanism would be hard to design for stimuli like RSS, where the sound levels are both positive and negative relative to the reference and the effects are both inhibitory and excitatory. Second, in three type IV neurons it was possible to obtain complete data samples with the 12- and 3-dB contrasts presented as separate stimulus sets and then
again with the 12- and 3-dB stimuli interleaved. Although some effects of the interleaving were seen (Reiss 2005), they were small and not consistently in the direction predicted by an adaptation mechanism.

It is also unlikely that a moderately fast adaptation process operating within a single stimulus presentation adjusts the gain for that presentation. In this case, a consistent decrease in weight amplitude should occur through the duration of the stimulus for the 12-dB contrast. Weights were computed in successive bins of 40–100 ms in 21 neurons (from this and previous studies) for which there were sufficient data to allow estimation of weights in short time windows. Again, changes in weights did occur, but they were not systematic and could not provide an explanation for the difference between 12- and 3-dB SD stimuli. Thus any adaptation process would have to operate on a very short timescale, <40 ms.

Based on the arguments of the preceding two paragraphs, it seems likely that the changes in weight magnitude are caused by a fast (essentially instantaneous, given the limits of the data available here) process like cochlear compression augmented by an additional fast process in the cochlear nucleus. We argued against an effect that can be represented by a single static nonlinearity that applies uniformly across frequency with the data in Fig. 6. However, in another paper (Bandyopadhyay et al., unpublished observations), we show that a frequency-dependent static nonlinearity, meaning a level-dependent gain mechanism that has a different shape in each frequency bin, can account for the changes in weight size with contrast. In the level-dependent model, the weights change with contrast because the quadratic model is approximating a nonlinear function in each frequency bin; the quadratic model’s gain is a compromise or average of many slopes because the true input-output function is nonlinear and usually saturating or non-monotonic. Thus the quadratic model’s weight is expected to be smaller when the stimulus contrast is large enough to extend the averaging over multiple different slopes.

This mechanism can also account for the poor prediction performance of the quadratic model at 12-dB spectral contrast. Because of the averaging process described in the preceding paragraph, the gain in each frequency bin does not correctly capture the actual nonlinear gain of the neuron and thus cannot accurately predict responses. One source of nonlinear gain adjustment is cochlear compression, but there must be additional mechanisms in the cochlear nucleus, as discussed earlier. There can be gain adjustment in the cochlear nucleus while retaining responses that are well fit by the quadratic model, as shown by the example of the VCN (Yu 2003). Chopper neurons of the VCN have first-order weights that are about twice as large as those of AN fibers but the quadratic model accurately predicts their responses. Thus the poor prediction shown here must be an effect of the circuitry specifically of the DCN.

It seems likely that inhibition of type IV neurons, especially by DCN type II neurons (Young and Davis 2002), can account for both the nonlinearities of the level-dependent weights and the poor prediction performance for 12-dB contrast. The importance of inhibitory inputs in shaping the responses of DCN neurons has been demonstrated (Davis et al. 1996; Nelken and Young 1994; Reiss and Young 2005; Spirou and Young 1989). Most important for this discussion, the BFIs of the excitatory and inhibitory inputs to DCN principal cells differ, so that the nonlinearities caused by inhibition differ in different frequency bands. It has also been shown that nonlinearity in the responses of DCN type IV neurons correlates well with stimuli and stimulus levels that produce responses in type II inhibitory neurons (Nelken and Young 1997; Nelken et al. 1997).

Note that even though the response predictions for 12-dB contrast stimuli were poor in type IV neurons, the weight functions obtained at 12-dB contrast were similar in shape to those obtained at 3-dB contrast where prediction is good (Fig. 4, C and E). Thus even though the receptive field derived from the 12-dB-contrast stimuli does not predict responses, it does have a basically correct shape and does provide information about the nature of spectral integration in the neuron in the absence of the nonlinearities that occur over larger ranges of stimulus level.

In conclusion, the results of this study suggest that both STRFs and spectral receptive fields depend on stimulus contrast. Furthermore, the results raise the possibility, supported by the results using level-dependent weight functions (Bandyopadhyay et al., unpublished observations), that STRFs derived from stimuli with “natural” contrasts on the scale of 12-dB SD may in reality represent averaged or “washed out” estimates of highly nonlinear, level- and frequency-dependent functions; this could partly account for why predictions from STRFs are often poor.

In general, decreasing spectral contrast will improve the model fit and the estimate of local curvature, and may be a better approach to studying the frequency selectivity of non-linear neurons, even though the stimulus statistics may not exactly resemble those of natural stimuli.

**Acknowledgments**

We thank S. Chase and B. May for comments on the manuscript. The authors acknowledge A. Fishbach, M. Anderson, S. Soujhiou, and T. Ropp for help in experiments and I. Bruce for providing the latest version of a proprietary cochlear model.

Present address for L. Reiss: Department of Speech Pathology and Audiology, University of Iowa, Iowa City, IA 52242.

**Grants**

This research was supported by National Institute on Deafness and Other Communication Disorders Grants DC-00115, DC-05211, and DC-00441.

**References**


