Published in final edited form as:

Neuroscience. 2010 January 20; 165(2): 612–620. doi:10.1016/j.neuroscience.2009.10.034.

Nonlinearity of Coding in Primary Auditory Cortex of the Awake Ferret

B. Shechter^{1,3} and **D.A. Depireux**^{2,3}

¹Department of Biology, University of Maryland, College Park

²Institute for Systems Research, University of Maryland, College Park

³Department of Anatomy and Neurobiology, School of Medicine, University of Maryland, Baltimore

Abstract

Neural computation in sensory systems is often modeled as a linear system. This first order approximation is computed by reverse correlating a stimulus with the spike train it evokes. The spectro-temporal receptive field (STRF) is a generalization of this procedure which characterizes processing in the auditory pathway in both frequency and time. While the STRF performs well in predicting the overall course of the response to a novel stimulus, it is unable to account for aspects of the neural output which are inherently nonlinear (e.g. discrete events and non-negative spike rates). We measured the STRFs of neurons in the primary auditory cortex (AI) of the awake ferret using spectro-temporally modulated auditory gratings, or ripples. We quantified the degree of nonlinearity of these neurons by comparing their responses to the responses predicted from their respective STRFs. The responses of most cells in AI exhibited a squaring, nonlinear relation to the stimuli used to evoke them. Thus, the nonlinearity of these cells was nontrivial, i.e. it was not solely the result of spike rate rectification or saturation. By modeling the nonlinearity as a polynomial static output function, the predictive power of the STRF was significantly improved.

Keywords

Auditory Cortex; Spectro-Temporal Receptive Field; Nonlinear coding; Broadband Sounds

Introduction

Neural processing of the auditory environment is an inherently nonlinear process. For instance, speech intelligibility decreases nonlinearly with reductions in signal to noise ratios or modulation depth (Chi et al. 1999; Deller JR. et al. 2000; Drullman et al. 1994). In a noisy environment, a healthy listener can easily attend a specific stream and perceive its content. However, there is a heightened sensitivity to masking following bilateral auditory cortex strokes (Tramo et al. 1990; Tramo et al. 2002). Patients with damage to AI describe having difficulty extracting low-contrast features. These studies implicate the role of higher stations

© 2009 IBRO. Published by Elsevier Ltd. All rights reserved.

Contact Information. Didier A Depireux, Institute for Systems Research, University of Maryland, College Park, College Park, MD 20742, depireux@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of the auditory pathway, including AI, in signal detection and enhancement, and their processing likely reflects the nonlinear aspects of competing sound envelopes, maskers and simple envelope contrast. Thus, the intelligibility of speech in noise is dependent on contrast between the spectral peaks and the background, and a complex spectral enhancement is key for a sound analysis of the auditory scene.

Few studies have characterized in the awake animal the nonlinear processing components involved in the analysis of competing sound envelopes and the relevant neural circuitries. In both anesthetized and awake animals, phase-locked responses to the sound envelope can deviate from linear predictors; these are considered nonlinearities, and may arise from both trivial and non-trivial sources. Even from purely physical constraints on neurons, we expect nonlinear responses. We define response nonlinearity that arises due to physical properties of neurons to be trivial. Such sources might include membrane time constants and synaptic factors. At the very least, there are absolute bounds imposed on the neural response by metabolic processes. The firing rate of a neuron has a maximum possible value, or saturation rate, and inhibition of its firing cannot induce a negative firing rate. While these nonlinearities are trivial in that they are expected, they still contribute to the essential processing of the auditory scene.

In addition to our expectation of trivial nonlinearities, there is neurophysiological evidence that more complex output processing occurs in cortex beyond that of thresholding and saturation, such as circuit convergence and delay patterns (Fishbach et al. 2003; Schnupp et al. 2001). Such nonlinearities, which arise due to network properties, and not purely physical properties of the neurons, are nontrivial nonlinearities. For instance, the average firing rates in cortex drop in magnitude from earlier stages in the pathway. An input-output nonlinearity at such a stage would enable a nonlinear compression scheme whereby essential information of the auditory scene is preserved—perhaps even enhanced—and features of lesser relevance are degraded or discarded. When trivial nonlinearities are included in models used to predict A1 responses to complex sound, the prediction success can be improved (Shechter and Depireux 2007); however these do not sufficiently describe the output nonlinearity (Machens et al. 2004; Young 2008).

At the level of cortex, processing is believed to be well *approximated* by linear systems, but having a nontrivial nonlinear component to responses (Ahmed et al. 2006; Calhoun and Schreiner 1998; Machens et al. 2004; Sahani and Linden 2002; Schnupp et al. 2001; Sen et al. 2001; Theunissen et al. 2000). Neurons in auditory cortex are tuned to extract specific aspects of the spectro-temporal content of sounds. The spectral profile encoded by these neurons is built up from features extracted in subcortical input neurons. The cortical representation of spectro-temporally rich, well parameterized sounds has been the subject of intense research (Depireux et al. 2001; Kowalski et al. 1996a; b; Linden et al. 2003; Schreiner and Calhoun 1994; Versnel et al. 1995). While details vary considerably between methods, auditory cortical processing can be modeled by linear systems which transform an input sound stimulus to a predicted neural response. These linear spectro-temporal receptive fields (STRFs) are useful as a first approximation in that they generally capture the overall trend of the response. However, by definition, they cannot provide a construct to characterize nonlinear processing. A well established method of obtaining a preliminary characterization of the output nonlinearity is to relate the neural response to its linear receptive field by a static input/output function (Lee and Schetzen 1965; Marmarelis 1993).

There is debate about the degree to which AI responses are nonlinear (Ahmed et al. 2006; Ahrens et al. 2008; Kowalski et al. 1996b; Machens et al. 2004; Schnupp et al. 2001; Theunissen et al. 2000; Valentine and Eggermont 2004); however, few studies have been carried out in the awake animal where, among others, recurrent projections to and from

subcortical structures and descending influence from higher cortical areas such as prefrontal cortex would impact processing in AI (Fritz et al. 2007; Knight et al. 1999).

In this study, we examine and classify the static output nonlinearities observed in cortical responses. Are these nonlinearities trivial in the sense that they are predetermined by physical constraints of the individual neurons? Or do they perform a computation on the input thereby enhancing features important to processing the auditory scene. Is a static output nonlinearity (either trivial or nontrivial) sufficient to account for the complex output processing of neurons in primary auditory cortex? To address these questions, we quantify the degree of linearity of responses in AI. We show that these neurons employ nontrivial nonlinearity in their computations, which reflects the extraction of spectrotemporal contrast.

Experimental Procedures

Recordings were made from 3 awake, 3 to 12 month old, domestic ferrets (*Mustela putorius furo*) surgically implanted with chronic multi-electrode arrays. A detailed description of recording and surgical procedures can be found in (Dobbins et al. 2007). Briefly, ferrets were anesthetized with Isoflurane (3% induction, 1.75% maintenance adjusted to keep heart rate, end-tidal CO₂ and SpO₂ within limits). Body temperature was maintained at 37.5°C with a feedback heating pad. The scalp was incised rostral-caudally along the midline from bregma to the nuchal crest. The temporalis muscles were partially resected, and 8 stainless steel screws were inserted around the skull to anchor the subsequent headpost, custom multi-electrode microdrive, and dental cement used to fix the experimental apparatus. A headpost was positioned rostrally over the skull. A craniotomy was made over the left AI, and to help prevent regrowth, the mitotic inhibitor 5-fluorouracil (5-FU) was applied (Spinks et al. 2003). The microdrive was slowly lowered into the craniotomy, and affixed with dental cement. Prior to surgery, the microdrive had been loaded with 6 to 12 independently adjustable micro-electrodes (12Drive-H, Neuralynx, Tucson AZ), with a custom made piece, resting on the edges of the craniotomy, to lower the electrodes in a honeycomb pattern over AI in such a way that the minimum distance between adjacent electrodes was 225µm. The experimental apparatus and headpost were fixed in place with additional dental cement. After surgery, the ferrets were given Banamine (1mg/kg) and Baytril (.2 mg/kg) for three recovery days before recording began. All surgical and experimental procedures were approved by the University of Maryland Animal Care and Use Committee and were in accord with NIH Guidelines on the care and use of laboratory animals.

Neural Recordings

Recording sessions took place inside a double-walled sound booth (IAC, Bronx, NY, Noise Isolation Class of 70dB) with no source of extraneous noise inside. The ferret was placed in a comfortable holder, with its head fixed using the implanted headpost to ensure the animal stayed within the calibrated sound field and to minimize movement noises. The animal was monitored through a closed-circuit video, and treats (e.g., Ferretone) were given between stimulus presentations to preserve wakefulness as deemed needed from the emergence of monitored slow-wave activity.

Recording sessions typically lasted 4–6 hours. Neural activity was recorded with parylene-coated tungsten microelectrodes (initial impedance 3–6 MΩ at 1 kHz, Micro Probe, Inc, Gaithersburg, MD). Electrodes were individually advanced (78µm) by manually turning a screw at the end of each recording session in preparation for the next day. All recordings were marked with the depth relative to when spikes were first detected upon lowering the electrode bundle. Spike events were obtained by a simple level-crossing. The signal was bandpass filtered with low and high cutoff frequencies of 300Hz and 3kHz, respectively. Spike times were assigned to the peak of the waveform with a resolution of 125 µsec. After

a recording session, spikes were sorted into classes using a modification of the MClust package (MClust, A. D. Redish), with the automated cluster cutter KlustaKwik (Harris and Redish 2002) which uses a CEM algorithm (*Conditional Expectation Maximization*) for which we used the Fourier transform, first and second principal components, energy, area, peak and valley of each event to classify spikes.

Stimulus Generation and Sound Presentation

All stimuli were generated digitally in MATLAB (Natick, MA), with custom software written by us, converted to an analog voltage (TDT RX6, Tucker David-Technologies, Alachua, FL) at 100kHz sampling rate, processed with an analog attenuator (TDT PA5), amplified (Crown DX-70) and then presented free field from an overhead speaker (Manger Transducer, Manger, Germany) located 1m at zenith relative to the animal's head. The sound field was calibrated to obtain a flat response (to within 1.5dB) from the loudspeaker where the animal's head would be located.

STRF Computation

Cells were characterized by measuring their spectro-temporal receptive field using a set of well structured, spectro-temporally rich broadband stimuli. These stimuli, known as auditory gratings—or ripples, each have a spectro-temporal profile modulated sinusoidally in spectrum and in time. The modulation of a grating is characterized in spectrum by its spectral density Ω (cycles/octave), in time by its angular frequency w (Hz), and in amplitude by its excursions away from the mean level of the stimulus (modulation depth ΔA , % of mean). The amplitude $S(x, t)$, corresponding to frequency f , where $x = \log_2(f/f_0)$ and f_0 the lower edge of the spectrum, is then adjusted at time t as

$$S(x, t) = L[1 + \Delta A \cdot \cos(2\pi(\Omega \cdot x + w \cdot t) + \phi)], \quad (1)$$

for a linear modulation. S defines the spectro-temporal envelope of the stimulus. L is the average level of the stimulus (measured as the RMS of the stimulus with reference to a 1kHz sine tone) and ϕ is the starting phase of the grating. Since the 100 tones per octave that make up a grating are logarithmically spaced, it does not elicit the perception of a pitch.

The spectral and angular densities of the stimulus set ranged from -5 to 5 cyc/oct and from 4 to 48 Hz, respectively. Each stimulus had a duration of 3 seconds and was presented with at least 4 repetitions at 65 dB SPL and at 90% modulation. We found this sound level optimal for eliciting activity on most electrodes. Because the base modulation rate was 4Hz, every stimulus was periodic with a base periodicity of at most 250msec, and every repetition yielded 12 measurements.

The derivation of the steady state STRF from grating stimuli is described in detail in (Klein et al. 2000). Briefly, a reverse correlation with the stimulus spectro-temporal envelope is used to obtain the STRF from the spike train elicited by the stimulus. The responses were binned into time windows as determined by the periodicity of the stimulus set. Bin width was chosen such that the maximal angular frequency was below the Nyquist frequency (from 10–15 msec, depending on w). All further spike rates for a particular cell used the same time resolution. In order to allow the response to reach a steady-state, the analysis was started at 250 msec post stimulus onset, thereby removing the effect of level transients present at the onset of the stimulus. We also measured the *evoked spike rate* r_0 elicited by the mean level of the stimulus (away from which the spectro-temporal envelope is modulated). The STRF measures the neural response to the stimulus modulations, whereas the evoked spike rate measures the response to the noise carrier, i.e., to a broadband noise of

equivalent level as the grating stimulus. These measures are in contrast to the spontaneous rate of firing, which exists in the absence of any stimulus.

SNR Computation

To determine the reliability of the STRF measured for the steady state response, we computed a signal-to-noise ratio (SNR) of the corresponding modulation transfer function (MTF, the 2-dimensional Fourier transform of the STRF). The MTF is the dual representation of the STRF, where each point $MTF(\Omega, w)$ is a complex number which represents the response of the neuron to an auditory grating with spectral density Ω (cycles/octave) and angular frequency w (Hz), see Eq.1. The amplitude of the complex number is the corresponding amplitude of the response and the phase is the corresponding phase lag between stimulus and response. 50 bootstrap estimates $\psi(\Omega, w)$ were generated for each point of the MTF by choosing with repetition from the entire set of responses obtained. The SNR of each point ($SNR_{\Omega, w}$) was taken to be the average power divided by the variance of the estimates. The SNR of the STRF was computed as the power-weighted mean of the SNR at each Ω, w point. $P_{\Omega, w}$ is the power of the MTF at the point (Ω, w) .

$$SNR_{\Omega, w} = \frac{\left| \sum_{bootstraps} \psi(\Omega, w) \right|^2}{N_{bootstraps} \cdot \sigma_{\psi}^2} \quad (2)$$

$$SNR = \frac{\sum_{\Omega, w} P_{\Omega, w} \cdot SNR_{\Omega, w}}{\sum_{\Omega, w} P_{\Omega, w}} \quad (3)$$

For a cell to be included in the subsequent analyses, we required that its SNR be greater than 1. Computing 50, 75, and 100 bootstrap estimates for a subset of recordings resulted in statistically equivalent SNRs; we therefore chose to compute SNR using 50 bootstrap estimates throughout.

Computation of Static Output Nonlinearities

The STRF can be used to predict responses to novel stimuli with complex spectro-temporal envelopes. This is achieved by convolving in time a spectro-temporal envelope with the STRF, and summing across all frequency bands. A predicted response spike rate $r(t)$ to a stimulus with spectro-temporal envelope $S(x, t)$ is therefore computed as

$$r_{lin}(t) = \iint STRF(x, \tau) \cdot S(x, t - \tau) d\tau dx. \quad (4)$$

Using the derived receptive fields, we computed the predicted responses to the stimuli used in measurement of the STRF (Fig. 1). We characterized the relationship between predicted and actual response by means of a parameterized fit.

In order to measure the *degree of linearity* in the response, we measured the power α of the relationship between predicted response and actual response: $r_{act}(t) \propto r_{lin}(t)^\alpha$. The exponent α was found by finding the slope of a linear fit *on a logarithmic scale*. Specifically, the evoked spike rate r_0 was added to the predicted response; the logarithm of this sum and the logarithm of the actual response were calculated for positive values; and the least squares linear fit was computed between both logarithms. The slope of this fit was α . A value of 1

implies a linear relationship between stimulus and response, whereas values different than 1 implies the existence of higher order processing.

α enabled us to measure the degree of linearity in auditory responses with a single quantity. Good as a first approximation, it provided a concise descriptor for comparing all recorded cells. After measuring the degree of linearity with α , we proceeded to characterize in more detail the functional relationship between the linear prediction and the recorded responses. Visual inspection of our entire dataset (as shown in Fig. 2, Fig. 3) showed no rate saturation, as would occur from the neurons reaching a maximal firing rate. Rate saturation would be indicated by a leveling-off of actual response when plotted as a function of the linearly predicted response (similar to a sigmoid function). Given that we did not observe saturation in any of our recordings, (even though it is usually found when measuring rate-level functions in the onset response to different kind of stimuli, such as pure tones (Heil 1997)), and because of the highly general quantitative power offered by polynomial approximations, we then computed the static output nonlinearity as the 3rd order polynomial P_{NL} which minimized the squared error between the predicted response and the actual response (Fig. 1). This allowed us to characterize the nonlinearity arbitrarily in terms of a power series, without over-emphasizing lower spike rates.

$$r_{poly}(t) = P_{NL}[r_{lin}(t)]. \quad (5)$$

In order to measure the efficacy of the computed static output nonlinearity, we recorded responses to new test stimuli. These stimuli, called temporally orthogonal ripple combinations (TORCs), had complex spectro-temporal envelopes defined as the sum of multiple auditory gratings (Klein et al. 2000). Each of the grating components comprising a TORC had the same spectral density and depth, but differed in angular frequency. Their initial phases ϕ were “optimized”, in the sense that their values were randomly selected so as to minimize the “peakiness” of the envelope, as measured by the peak to RMS ratio. The STRF was used to compute the linear predicted response r_{lin} , and in conjunction with P_{NL} , also the prediction r_{poly} . We measured the RMS error between predictions and actual responses obtained.

Results

The results presented are taken from a corpus of 506 cells (meeting the requirement of SNR > 1) recorded in primary auditory cortex (AI) of 3 animals. Each animal was recorded from 5 days a week over the course of 6 weeks. Figure 2 and Figure 3 demonstrate 10 example cells for which the analyses described above were computed. The STRF of each cell is depicted alongside the relationship observed between the linear predicted responses and the actual responses.

Quantifying the Degree of Linearity

We measured the power (α) of the regression between predicted response and actual response by computing the slope of a linear regression analysis performed on a logarithmic scale. This value characterizes the degree of linearity these neurons showed in their responses. The data shown in Fig. 2 were recorded from neurons which responded linearly, with the exception of a trivial rectifying nonlinearity ($\alpha \approx 1$). Those in Fig. 3 responded with nontrivial output nonlinearities ($\alpha > 1$).

Across all cells we recorded in AI, the mean of all measured α 's was 1.92 ± 0.76 (Fig. 4). Thus, the relationship between predicted linear response and actual response was nontrivial for most cells, with a power close to 2 (squaring).

Characterization of the Nonlinearity

We next quantified the relation between the predicted output from the STRF and the actual responses recorded for each neuron. A 3rd order polynomial fit P_{NL} was obtained which minimized the squared error between predicted and actual responses. Using a χ^2 goodness-of-fit test, we determined that this polynomial quantified the observed nonlinearity accurately ($p < 0.001$ for each cell). To obtain a *predicted* response, we evaluated the polynomial P_{NL} on the output of the STRF (Eqs. 4,5). In order to determine the utility of P_{NL} , we quantified the error when used to predict responses to *new sounds* (i.e., not those used to construct the STRF and P_{NL}). With respect to the TORC stimuli, we measured the RMS error between both linear and nonlinear predictions and the actual responses obtained during the recording session. In Fig. 5, we show the TORC stimuli prediction errors computed for the 6 nonlinear cells shown earlier in Fig. 3. The black dotted curves represent the errors from a linear prediction only, whereas the gray curves represent the errors computed by applying a polynomial output nonlinearity onto the linear prediction, i.e. the linear prediction was first computed and used as input to the polynomial nonlinearity. Note that not only are the RMS errors reduced, but the gray curves are narrower in comparison to their respective black curves. This corresponds to a reduction in the variability of the predicted values. In Fig. 6, we demonstrate the improved predictive power of the STRF when using a static output nonlinearity for the population of cells in this study. We show a significant decrease ($p < 0.001$, Wilcoxon signed-rank) in both RMS error and variance of error when using the linear predictor (STRF) and an output nonlinearity, as opposed to the linear prediction alone (Fig. 6).

In order to assess the importance of the higher-order, non-trivial component of the nonlinearity, we compared the prediction error obtained when using the polynomial static output nonlinearity to the linear prediction after addition of the evoked spike rate and half-wave rectification. We found great similarity between the reduction in variance of error obtained by using either the linear prediction or its half-wave rectification as baselines (Fig. 6,7). We also measured the change in correlation (as measured by the correlation coefficient ρ) between the predicted and actual responses. There was a significant increase (5%, $p < 0.001$, Wilcoxon signed-rank) in correlation with half-wave rectification of the linear prediction (Fig. 7, middle). There was a further, more striking, increase (8%, $p < 0.001$, Wilcoxon signed-rank) in correlation when using the polynomial static output nonlinearity, as opposed to half-wave rectification of the linear response (Fig. 7, right).

Discussion

We have demonstrated that most cells in primary auditory cortex of the awake ferret respond with a quantifiable degree of nonlinearity to the spectro-temporal envelope of complex sounds. Furthermore, this nonlinearity—a squaring nonlinearity—is not solely a result of response rate saturation and half-wave rectification. Previous published characterization of STRFs in cortex studies have often limited themselves to strict linear models, with the possible exception of the initial studies in the early 1980s by Eggermont et al. (1983). More recent studies have used nonlinearities for specific purposes, mentioned below.

Significance of $\alpha \approx 2$ and the implication of squaring nonlinearities

That $\alpha \approx 2$ implies not only that responses of cortical neurons are not linear, but that their nonlinearities are nontrivial. The range of exponents seen in our sample of AI cells may reflect a range suitable for nonlinear computations that can be implemented within the limitation metabolically imposed by a maximum firing rate. The mean value of 2 is not so high that the output nonlinearity is just an over-rectification, with extreme sensitivity to small changes in input. Such a nonlinearity would have a high firing threshold and high

gain, effectively reducing the input range cells can encode. Given the overall reduction in firing rate in cortex when compared to lower auditory centers, these nontrivial nonlinearities suggest cortical neurons are using an encoding scheme which enhances specific features.

What computations can be achieved by this static nonlinearity? First, a squared output can encode a phase-specific component of Fourier energy in a local region of the stimulus. This may be an ideal way of preserving both local amplitude and phase information (Pollen and Ronner 1982). Second, squaring enhances stimulus selectivity: tuning of cells for stimulus parameters becomes narrower, and its edges steeper (see Fig.1 Predictions). Third, and probably most relevant, the squared outputs of linear filters are the building blocks for an energy model: combining the squared output of neighboring cells, all encoding the same region of the tonotopic axis and tuned to the same range of spectral densities and temporal periodicities, but differing in phase, results in responses whose characteristics are the auditory equivalent of the complex cell responses of visual cortex. In the standard model of building a complex V1 cell from the output of simple cells, the squared outputs of linear cells, which respond to subregions of the responsive region of the complex cell, are added together. Since neighboring simple cells in V1 tend to have receptive fields in quadrature (Pollen and Ronner 1981), the resulting sum has the characteristics needed to generate the response properties of complex cells (Ohzawa et al. 1990; Rust et al. 2005; Watson and Ahumada 1985).

Exponential vs. Polynomial output nonlinearity

The exponent α provides a compact way of characterizing the degree of linearity present in the neural response. The simplicity in using a single number allows for an efficient cross-cell comparison of the entire population of responses. Nevertheless, the complexity of the input-output nonlinearity is not fully characterized by a single number. The exponential is highly restrictive in the range of functions it can approximate. Therefore, after performing an overall characterization, we quantified the specific nonlinearity for each cell with a 3rd order polynomial function. The use of polynomials allows a broad range of nonlinear functions to be captured, while also providing an efficient parameterization using a small number of coefficients. Given that no responses showed spike rate saturation, the static output nonlinearities were well fit by a polynomial.

Overall static output nonlinearity vs. a per channel nonlinearity

The introduction of the static output nonlinearity is to a first approximation a global measure of the output characteristics of a cortical neuron: the simplest nonlinearity is a half-wave rectifier, as dictated by the physical limitations of neurons. Our model consists in a linear model (the STRF) followed by an instantaneous non-linearity. The STRF models the computations carried out by the auditory pathway up to and including the neuron. This information is then conveyed as firing output in a manner modeled to a large extent by the static output nonlinearity. In our model of a linear system followed by an instantaneous nonlinearity, the physiological properties of the cortical neuron result in the nonlinearity, and become an *explicit* feature of the neural computation. This is not to imply that in the physical instantiation of our model, the STRF is a representation of *all the neural processes* up until that point, nor that the static output nonlinearity has to do *exclusively* with the physical properties of the neuron producing a spike. However, in determining the algorithms used by the auditory pathway, the output nonlinearity is essential to the neural computation and its associated properties of noise-robustness and efficiency, which we believe are typical of cortical representations.

Each neuron we recorded from was actually receiving inputs from neurons with a range of best frequencies. The contribution from each of its spectrally distributed inputs should not

necessarily be assumed linear in modulation. Ahrens et al. (2008) propose an overall input nonlinearity which conditions the input to the STRF based on stimulus level. One could extend the concept of input nonlinearity further to condition the input to the cortical model in a way that accounts for the subcortical processing—not just limited stimulus level. Perhaps, a more realistic extension of our model would entail the neuron integrating each spectral band nonlinearly before contributing the overall firing output. That is, instead of only a **global** static output nonlinearity, each frequency band should have its own **local** static nonlinearity. This more closely models the flow of information from downstream neurons and its subsequent integration for further processing. However, the main difficulty is that we have access only to the temporal output of the neurons, a one dimensional quantity. The use of our broadband sounds for spectro-temporal characterization is sufficient in the case of a global static output nonlinearity, but may not allow the disambiguation of many local nonlinearities.

Phase locked vs. phase independent nonlinearity

An established measure of complex (nonlinear) tuning is f_1/f_0 (Ahmed et al. 2006; Hubel and Wiesel 1965; Movshon et al. 1978a; b). This ratio measures the degree to which a neuron is phase locked (f_1) to modulations in the stimulus relative to the degree to which the spike rate is elevated irrespective of modulation phase (f_0). This measure derives from the study of complex visual cortex neurons, which respond to a grating with an elevated firing rate throughout its presentation, but without specificity for any stimulus phase. Such a phase independent nonlinearity cannot be measured reliably in neurons which have low evoked firing rates and an appreciable response locked to stimulus modulations. Trivial output nonlinearities, namely rectification at 0 spikes/sec, are likely to interfere with the measurement of the evoked firing rate during stimulus presentation. Such rectification would be measured as an artificial elevation of the evoked rate and a decrease in the magnitude of the phase locked response. Despite the difficulty in measuring phase independent complexities in the response of AI neurons, it is plausible that such rate encodings of the sound environment co-exist with spike timing based schemes and may serve to extract distinct components from the sound envelope (Malone et al. 2007; Zheng and Escabi 2008).

Inherent response variability

The predictive power of a polynomial static output nonlinearity used in conjunction with the linear receptive field was demonstrated by an overall reduction in the prediction error. Despite this improvement, the response characterization remains deterministic, in which a specific stimulus spectro-temporal envelope is predicted to evoke a unique response. However, in multiple presentations of a single stimulus, the responses of single neurons are highly variable from one trial to the next. The responses from two such neurons are displayed in Fig. 8. Throughout stimulus presentation, even when the same response is predicted, the actual response takes on multiple values. A static output nonlinearity, such as is described in this paper, is ill equipped to account for these statistics. Only a formulation which takes other state variables into account and is probabilistic in nature can robustly characterize the response and its inherent variability.

Dynamic tuning/adaptive tuning

The analysis in this study was conducted only after responses were allowed to reach a steady state regime. However, there is ample evidence that tuning at the level of primary auditory cortex evolves dynamically during stimulus presentation (Shechter et al. 2009). If the input-output properties of auditory neurons continuously evolve, the current static methods of characterization would be inadequate. The STRF method utilizes a short window of stimulus

history to predict an instantaneous response, but its features (used to make that prediction) are *independent* of the stimulus and its history.

In conclusion, the majority of cells found in AI of the awake ferret responded nonlinearly with respect to the stimuli used to evoke them. This nonlinearity was nontrivial in that it had a squaring effect ($\alpha \approx 2$). Yet, despite the utility of static output nonlinearities (as measured by the improvement in predictive power), such a framework is insufficient to fully account for the measured nonlinearity in cortical responses.

Abbreviations

AI	primary auditory cortex
IC	inferior colliculus
MTF	modulation transfer function
RMS	root-mean-square
SNR	signal-to-noise
STRF	spectro-temporal receptive field
TORC	temporally orthogonal ripple combination

Acknowledgments

The authors thank Yadong "KK" Ji for extensive help in animal care and data acquisition, and the referees for very useful suggestions that improved the manuscript. This research was funded by NIH/NIDCD RO1 DC005937 awarded to DAD.

References

- Ahmed B, Garcia-Lazaro JA, Schnupp JW. Response linearity in primary auditory cortex of the ferret. *The Journal of physiology*. 2006; 572:763–773. [PubMed: 16497713]
- Ahrens MB, Linden JF, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J Neurosci*. 2008; 28:1929–1942. [PubMed: 18287509]
- Calhoun BM, Schreiner CE. Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. *European Journal of Neuroscience*. 1998; 10:926–940. [PubMed: 9753160]
- Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. *Journal of the Acoustical Society of America*. 1999; 106:2719–2732. [PubMed: 10573888]
- Deller, JR., JR; Hansen, JHL.; Proakis, JG. *Discrete-Time Processing of Speech Signals*. Hoboken, NJ: 2000.
- Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophys*. 2001; 85:1220–1234.
- Dobbins HD, Marvit P, Ji YD, Depireux DA. Chronically Recording with a Multi-Electrode Array Device in the Auditory Cortex of an Awake Ferret. *J Neurosci Meth*. 2007; 159
- Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America*. 1994; 95:1053–1064. [PubMed: 8132899]
- Eggermont JJ, Aertsen AM, Johannesma PI. Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field. *Hear Res*. 1983; 10:191–202. [PubMed: 6602800]

- Fishbach A, Yeshurun Y, Nelken I. Neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. *J Neurophysiol.* 2003; 90:3663–3678. [PubMed: 12944531]
- Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J Neurophysiol.* 2007; 98:2337–2346. [PubMed: 17699691]
- Harris K, Redish AD. *KlustaKwik and MClust-3.3.* 2002
- Heil P. Auditory cortical onset responses revisited. II. Response strength. *J Neurophysiol.* 1997; 77:2642–2660. [PubMed: 9163381]
- Hubel DH, Wiesel TN. Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *J Neurophysiol.* 1965; 28:229–289. [PubMed: 14283058]
- Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of Computational Neuroscience.* 2000; 9:85–111. [PubMed: 10946994]
- Knight RT, Staines WR, Swick D, Chao LL. Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta psychologica.* 1999; 101:159–178. [PubMed: 10344184]
- Kowalski N, Depireux DA, Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology.* 1996a; 76:3503–3523. [PubMed: 8930289]
- Kowalski N, Depireux DA, Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology.* 1996b; 76:3524–3534. [PubMed: 8930290]
- Lee YW, Schetzen M. Measurement of the Wiener kernels of a nonlinear system by cross-correlation. *Int J Control.* 1965; 2:237–254.
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *Journal of Neurophysiology.* 2003; 90:2660–2675. [PubMed: 12815016]
- Machens CK, Wehr MS, Zador AM. Linearity of cortical receptive fields measured with natural sounds. *Journal of Neuroscience.* 2004; 24:1089–1100. [PubMed: 14762127]
- Malone BJ, Scott BH, Semple MN. Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol.* 2007; 98:1451–1474. [PubMed: 17615123]
- Marmarelis, V. *Nonlinear dynamic modeling of physiological systems.* Wiley-IEEE Press; 1993.
- Movshon JA, Thompson ID, Tolhurst DJ. Receptive field organization of complex cells in the cat's striate cortex. *The Journal of physiology.* 1978a; 283:79–99. [PubMed: 722592]
- Movshon JA, Thompson ID, Tolhurst DJ. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of physiology.* 1978b; 283:53–77. [PubMed: 722589]
- Ohzawa I, DeAngelis GC, Freeman RD. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science (New York, NY).* 1990; 249:1037–1041.
- Pollen DA, Ronner SF. Phase relationships between adjacent simple cells in the visual cortex. *Science (New York, NY).* 1981; 212:1409–1411.
- Pollen DA, Ronner SF. Spatial computation performed by simple and complex cells in the visual cortex of the cat. *Vision Res.* 1982; 22:101–118. [PubMed: 7101735]
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque v1 receptive fields. *Neuron.* 2005; 46:945–956. [PubMed: 15953422]
- Sahani, M.; Linden, J. How linear are auditory cortical responses?. In: Becker, T.; Obermayer, editors. *Advances in Neural Information Processing Systems.* Vancouver: MIT press, MA; 2002. p. NS02
- Schnupp JW, Msrac-Flogel TD, King AJ. Linear processing of spatial cues in primary auditory cortex. *Nature.* 2001; 414:200–204. [PubMed: 11700557]
- Schreiner CE, Calhoun BM. Spectral envelope coding in cat primary auditory cortex: Properties of ripple transfer functions. *Aud Neurosci.* 1994; 1:39–61.
- Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *Journal of Neurophysiology.* 2001; 86:1445–1458. [PubMed: 11535690]
- Shechter B, Depireux DA. Stability of spectro-temporal tuning over several seconds in primary auditory cortex of the awake ferret. *Neuroscience.* 2007; 148:806–814. [PubMed: 17693032]

- Shechter B, Dobbins HD, Marvit P, Depireux DA. Dynamics of spectro-temporal tuning in primary auditory cortex of the awake ferret. *Hear Res.* 2009; 256:118–130. [PubMed: 19619629]
- Spinks RL, Baker SN, Jackson A, Khaw PT, Lemon RN. Problem of dural scarring in recording from awake, behaving monkeys: a solution using 5-fluorouracil. *J Neurophysiol.* 2003; 90:1324–1332. [PubMed: 12904511]
- Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience.* 2000; 20:2315–2331. [PubMed: 10704507]
- Tramo MJ, Bharucha JJ, Musiek FE. Music perception and cognition following bilateral lesions of auditory cortex. *Cognitive Neurosci.* 1990; 2:195–212.
- Tramo MJ, Shah GD, Braida LD. Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology.* 2002; 87:122–139. [PubMed: 11784735]
- Valentine PA, Eggermont JJ. Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. *Hear Res.* 2004; 196:119–133. [PubMed: 15464309]
- Versnel H, Kowalski N, Shamma SA. Ripple analysis in ferret primary auditory cortex: III. Topographic distribution of ripple response parameters. *Aud Neurosci.* 1995; 1:271–285.
- Watson AB, Ahumada AJ Jr. Model of human visual-motion sensing. *Journal of the Optical Society of America A-Optics & Image Science.* 1985; 2:322–341.
- Young ED. Neural representation of spectral and temporal information in speech. *Philosophical transactions of the Royal Society of London.* 2008; 363:923–945. [PubMed: 17827107]
- Zheng Y, Escabi MA. Distinct roles for onset and sustained activity in the neuronal code for temporal periodicity and acoustic envelope shape. *J Neurosci.* 2008; 28:14230–14244. [PubMed: 19109505]

Figure 1. Linear-Nonlinear STRF Model. Using the spectro-temporal receptive field (STRF) derived by reverse correlation, a linear predicted response rate is obtained by convolution of stimulus and STRF in time and summation across frequency. The nonlinear prediction is obtained by applying the polynomial static output nonlinearity function to the linearly predicted response. A squaring output nonlinearity can enhance stimulus selectivity, demonstrated by narrower peaks and steeper edges in the response.

Figure 2.

Linear Responses. The responses of four recorded neurons are depicted whose outputs were linear with respect to modulations in the stimulus spectro-temporal envelopes. Each point represents a point in time at which the actual response rate was sampled and a predicted response rate was computed. Values of α which were near 1 signified linearity. The STRF of each neuron is inset.

Figure 3.

Nonlinear Responses. The responses of six recorded neurons are depicted whose outputs were nonlinear with respect to modulations in the stimulus spectro-temporal envelopes. Each point represents a point in time at which the actual response rate was sampled and a predicted response rate was computed. Values of α which were different than 1 signified nonlinearity. For many cells recorded, the value of α was about 2, implying a squaring nonlinearity in their outputs. The STRF of each neuron is inset.

Distribution of Power Relationship between Linearly Predicted and Actual Responses

Figure 4.

Distribution of α . The degree of linearity α was measured by computing the slope of a linear fit between linearly predicted response and actual response on a logarithmic scale. Values of α near 1 demonstrate linearity, whereas values different than 1 demonstrate nonlinearity. For the population of cells recorded, the mean value of α was 1.92 ± 0.76 . This implies the existence of a squaring output nonlinearity for most cells in AI. 5 outliers not shown.

Figure 5.

Linear and Nonlinear Prediction Errors. For the 6 nonlinear cells depicted in Fig. 3, we plot the prediction errors (with respect to the actual response) when using only the STRF (black, dashed) and when using the STRF with static output nonlinearity (gray). There were reductions in both RMS prediction error and variance of prediction error for each neuron when using a static output nonlinearity.

Figure 6.

Reduction of Prediction Error with Static Output Nonlinearity. Distributions of the RMS prediction error (left) and variance of error (right) show a significant improvement in predictive power when using a static output nonlinearity, as opposed to the STRF alone ($p < 0.001$, Wilcoxon signed-rank). The graphs show the difference in prediction errors (with respect to the actual response) normalized to the linear prediction error. Negative values signify a reduction in the prediction error when using a static output nonlinearity. 4 and 8 outliers are not shown in the left, right distributions.

Figure 7.

Comparison of the Static Output Nonlinearity to Half-wave Rectification. **Left:** The distribution of change in variance of error (left) when using a static output nonlinearity, normalized to the error from half-wave rectification of the linear prediction, shows little difference when compared to the same measure using only the linear prediction (no half-wave rectification, Fig. 6). **Middle:** The correlation of the predicted response with respect to the actual response increased significantly when using a half-wave rectification compared to the linear prediction alone (5%, $p < 0.001$, Wilcoxon signed-rank). **Right:** However, there was a further, more striking increase in correlation when using the polynomial static output nonlinearity (8%, $p < 0.001$, Wilcoxon signed-rank) normalized to the correlation value of the half-wave rectified linear prediction. 3 and 6 outliers are not shown in the middle, right distributions.

Figure 8.

Inherently Variable Responses. The responses of two recorded neurons are depicted whose outputs were highly variable. Each point represents a point in time at which the actual response rate was sampled and a predicted response rate was computed. While an overall trend can still be seen, for any linearly predicted response value, there is a wide range of actual responses. For these neurons, while a static output nonlinearity will improve prediction errors, they will inevitably be high. The STRF of each neuron is inset.