Wiener-Volterra Characterization of Neurons in Primary Auditory Cortex Using Poisson-Distributed Impulse Train Inputs

Martin Pienkowski, Greg Shaw and Jos J. Eggermont
Department of Physiology and Biophysics, Department of Psychology
University of Calgary
Calgary, Alberta, Canada

Abstract

An extension of the Wiener-Volterra theory to a Poisson-distributed impulse train input was used to characterize the temporal response properties of neurons in primary auditory cortex (AI) of the ketamine-anesthetized cat. Both first- and second-order “Poisson-Wiener” (PW) models were tested on their predictions of temporal modulation transfer functions (tMTFs), which were derived from extracellular spike responses to periodic click trains with click repetition rates of 2 to 64 Hz. Second-order (i.e., nonlinear) PW fits to the measured tMTFs could be described as very good in a majority of cases (e.g., predictability ≥ 80%), and were almost always superior to first-order (i.e., linear) fits. In all sampled neurons, second-order PW kernels showed strong compressive nonlinearities (i.e., a depression of the impulse response), but never expansive nonlinearities (i.e., a facilitation of the impulse response). In neurons with lowpass tMTFs, the depression decayed exponentially with the interstimulus lag, whereas in neurons with bandpass tMTFs, the depression was typically double-peaked, and the second peak occurred at a lag that correlated with the neuron’s best modulation frequency. It appears that modulation-tuning in AI arises in part from an interplay of two nonlinear processes with distinct time courses.
Introduction

Neurons in primary auditory cortex (AI) of anesthetized animals typically respond only to the onsets of sounds whose rates of amplitude modulation (AM) exceed a few tens of Hz (or which are unmodulated), though this is not always the case in the awake condition (e.g., Recanzone, 2000; Lu et al., 2001; Wang et al., 2005). This strongly nonlinear behavior, potentiated under anesthesia, leads to an innate preference for slowly-modulated sounds, such as speech and other conspecific communication calls. It has recently been proposed that modulation-tuning across neural populations is optimized for discriminating between and within classes of natural sounds, including conspecific calls (Woolley et al., 2005). While the synchronized response of most AI neurons drops off with increasing AM rate in simple lowpass fashion, some neurons exhibit bandpass tendencies. The cellular mechanisms underlying modulation-tuning in AI, however, remain unclear.

The degree of neuronal synchronization to stimulus temporal fluctuations is typically quantified by computing the temporal modulation transfer function (tMTF) from responses to periodic or AM sounds (e.g., Eggermont, 1994; 2002; Lu and Wang, 2000). The tMTFs of auditory neurons have also been derived indirectly, via system identification methods (for a review see, e.g., Wu et al., 2006). Most of these latter efforts, however, relied on the assumption of linearity: the tMTF was derived, essentially, by Fourier transformation of the neuronal impulse response (e.g., Depireux et al., 2001; Theunissen et al., 2001; Woolley et al., 2005). While linear models have provided and continue to provide much insight into the nature of central auditory processing, it is a fact that AI neurons often exhibit profound nonlinearities (e.g., Bar-Yosef et al., 2002). Thus, it is not surprising that linear models are highly context-dependent (i.e., depend on the test-stimulus used in their estimation), and perform poorly when used to predict responses to complex stimuli (e.g., Sahani and Linden, 2003; Machens...
et al., 2004; Christianson et al., 2008). The addition of an ad-hoc, static nonlinear transformation to the linear model is enjoying a recent resurgence (e.g., Brenner et al., 2000; Nagel and Doupe, 2006; Atencio et al., 2008), and does improve predictions (Sharpee et al., 2008), but still leaves the temporal nonlinearities uncharacterized. The more comprehensive ad-hoc model of Ahrens et al. (2008), which explicitly includes temporal nonlinearities, arguably represents the present state-of-the-art in the modeling of the AI neuron response.

A rigorous and general nonparametric theory of system identification, capable, in principle, of characterizing the input-output relation of any (time-invariant) nonlinear system with memory, was pioneered long ago by Vito Volterra and Norbert Wiener (Volterra, 1930; Wiener, 1958; for an accessible introduction to the theory see Marmarelis and Marmarelis, 1978). In essence, the theory combines the Taylor series expansion with the convolution operator to capture the influence of a system’s memory, or of stimulus interactions, on the impulse response. Thus, the second-order Wiener-Volterra “kernel” effectively corrects the impulse response for interactions with single preceding impulses (as a function of the interval between the two impulses), the third-order kernel corrects for interactions with pairs of preceding impulses (as a function of the two intervals between the three impulses), and so on. One practical limitation of the theory, especially in its application to neural systems, is that the amount of data required to estimate kernels at a given signal-to-noise ratio (SNR) grows exponentially with the kernel order. Above second-order, the recording times to acquire sufficient numbers of spikes are very long, and the condition of response stationarity becomes very difficult to meet. Nevertheless, providing that second-order interactions are strong, the improvement of second-order over linear predictions can be substantial (e.g., Marmarelis and Naka, 1972; Juusola et al., 1995).
Here we applied what has been termed the “Poisson-Wiener” (PW) model (Marmarelis, 2004) to characterize the temporal response properties of AI neurons in the ketamine-anesthetized cat. The PW model represents a reformulation of the Gaussian white noise-based Wiener theory around a Poisson-distributed impulse train test-input (Krausz, 1975; Epping and Eggermont, 1986), which evokes a much more robust response in AI. Both first- and second-order PW models were tested on their predictions of neuronal tMTFs, which were derived from responses to periodic click trains at rates of 2 to 64 Hz. We found that second-order fits were substantially better than linear fits in almost all cases, and could be described as very good (predictability ≥ 80%) in a majority of cases.

Methods

Animal Preparation, Sound Stimulus Presentation, and Spike Recording

Animal use was approved (BI 2001–021) and reviewed on a yearly basis by the Life and Environmental Sciences Animal Care Committee of the University of Calgary, according to the guidelines set by the Canadian Council of Animal Care. Cats were deeply anesthetized with 25 mg/kg ketamine hydrochloride and 20 mg/kg sodium pentobarbital, injected intramuscularly. Lidocaine (20 mg/ml) was injected subcutaneously prior to the dissection of tissue, and craniotomy over the right temporal lobe. The dura was resected to expose AI, roughly bounded by the anterior and posterior ectosylvian sulci. A mixture of 0.2 ml acepromazine (0.25 mg/ml) and 0.8 ml atropine methyl nitrate (25 mg/ml) was periodically administered subcutaneously at 0.25 ml/kg body weight to reduce secretions into the respiratory tract. Small doses of ketamine were administered as required, typically every hour, to maintain a state of areflexive anesthesia (average across cats: 11.8 mg/[kg·h]; range: 8.8–16.7
All cats had normal hearing, as assessed using auditory brain stem response thresholds (Noreña and Eggermont, 2005).

Acoustic stimuli were presented in an anechoic room from a speaker system (Fostex RM765 in combination with a Realistic super-tweeter) that produced an approximately flat spectrum (± 5 dB up to 40 kHz, measured at the cat’s head). The speaker was placed about 30° from the midline in the contralateral field, ~50 cm from the cat’s left ear. Stimuli were generated in MATLAB and transferred to a sound-delivery system (Tucker Davis Technologies RP2.1).

The cat was secured with one screw cemented on the head without any other restraint. Two arrays of eight electrodes (Frederic Haer) each with impedances between 1 and 2 MΩ were used for spike acquisition. The depth of recording was between 700 and 1,200 µm, and thus the electrodes were likely in deep layer III or layer IV. The recorded potentials were amplified 10,000 times (Tucker Davis Technologies RA16PA), filtered between 0.3 and 3 kHz, and digitized by a multi-channel data-acquisition system (Tucker Davis Technologies RX5). Spikes were identified online using a single trigger set well outside of the noise floor, and sorted offline using an automated procedure based on principal component analysis and K-means clustering (Eggermont, 1990). Three of the best-sorted waveforms on an electrode, each potentially representing a single-unit, were combined for statistical purposes to form a multiple single-unit (MSU) spike train. We have previously shown that single-unit and MSU spike trains give nearly identical tMTFs (Eggermont, 1998).

The Poisson-Wiener Series

The output of a nonlinear “black box” system as a function of time, y(t), can be approximated by a series of orthogonal functionals (i.e., functions of a function), G_i, of the input to the sys-
tem, x(t), and its “Poisson-Wiener” (PW) kernels, p_i (after Krausz [1975] and Epping and Eggermont [1986]):

\[ y(t) = \sum_{i=0}^{\infty} G_i(p_i, x(s)), s \leq t \quad (1) \]

where,

\[ E(G_i \cdot G_j) = 0, \forall i \neq j \quad (2) \]

The first three functionals are:

\[ G_0(t) = p_0 \quad (3) \]

\[ G_1(t) = \int_{0}^{\infty} d\tau p_1(\tau)[x(t-\tau) - \lambda]; \quad \tau \geq 0 \quad (4) \]

\[ G_2(t) = \int_{0}^{\infty} d\tau \int_{0}^{\infty} d\theta p_2(\tau, \theta)[x(t-\tau) - \lambda][x(t-\tau-\theta) - \lambda]; \quad \tau \geq 0, \ \theta > 0 \quad (5) \]

where \( \lambda \) is the mean impulse rate of the Poisson-distributed test-input. Following Lee and Schetzen (1965), the PW kernels were estimated by cross-correlation of the system’s spike output with the Poisson-distributed impulse train input:

\[ p_0 = E[y(t)] \quad (6) \]

\[ p_i(\tau) = \lambda^{-1} E[y(t)x(t-\tau)] - p_0; \quad \tau \geq 0 \quad (7) \]

\[ p_2(\tau, \theta) = \lambda^{-2} E[y(t)x(t-\tau)x(t-\tau-\theta)] - p_1(\tau) - p_1(\tau + \theta) - p_0; \quad \tau \geq 0, \ \theta > 0 \quad (8) \]

where E denotes the expected (mean) value of the cross-correlation across time. The kernels may be qualitatively interpreted as follows. The zeroth-order kernel is equal to the average system activity (i.e., firing rate), whether spontaneous or stimulus-evoked. The first-order kernel represents the average response that is time-locked to an individual impulse, or in other words, the mean-rate-corrected peristimulus time-histogram (PSTH). The second-order kernel, in this “predictive form”, represents the average facilitation or depression of the response to an impulse, when that impulse is preceded by another impulse by \( \theta \) ms, regardless of intervening impulses. A second-order PW series gives an exact prediction of a system’s
response to any two-impulse experiment, or, if the system nonlinearity is at most quadratic, an exact prediction of the response to any experiment (within assumptions of stationarity, etc.).

First- and second-order PW kernels were computed by discretizing Equations 7 and 8, respectively, using a time bin of 1 ms. First-order kernels were subsequently smoothed with a 5-bin uniform filter, and second-order kernels with a 21x21-bin uniform filter. Stimuli for kernel estimation were 15 minute-long Poisson-distributed trains of acoustic clicks (abbreviated as “Poisson-clicks”), and also (for comparative purposes) Poisson-distributed trains of tone pips (abbreviated as “Poisson-pips”; different frequency pips were treated as identical “impulses”). All individual clicks and pips were presented at a peak level of 65 dB pe SPL. Clicks were initially generated at the mean rate of 8 /s, which was subsequently reduced to ~6.7 /s (λ in Equations 3–8) by imposing a minimum dead time of 20 ms between successive clicks, in order to increase the evoked neural firing rate. Pips were randomly and independently drawn from a pool of 81 frequencies that spanned five octaves (0.3–10 kHz, or 0.6–20 kHz, or 1.2–40 kHz, depending on responses to a previous search stimulus) in equal logarithmic steps. Each frequency was presented at an average rate of 1 per 4 s so the average aggregate pip presentation rate was 20 /s. The time course of each pip was shaped by a gamma-function:

\[
\gamma(t) = (t/4)^2 \exp(-t/4); \quad 0 \leq t \leq 50\text{ms} \quad (9)
\]

The gamma-envelope is above half-maximum power (i.e., is within 6 dB of the peak level) over approximately 3 to 17 ms post pip-onset, and decays by about 64 dB at 50 ms, at which point the pip was truncated.
**Frequency Tuning Curves**

MSU frequency tuning was assessed on the basis of responses to a pseudorandom sequence of tone pips, differing in frequency and intensity, which were presented one at a time at the rate of 4 /s. The same five octaves were spanned by 27 frequencies in equal logarithmic steps. Each frequency was presented at eight peak intensity levels, from -5 to 65 dB SPL, in 10 dB steps. Every frequency-intensity combination was repeated 10 times in the random sequence.

**Spectrotemporal Receptive Fields (STRFs)**

STRFs were computed from responses to the 15 minute-long Poisson-pips stimulus. Our procedure for computing STRFs was first detailed in Valentine and Eggermont (2004) and updated in Eggermont (2006). Briefly, each spike registered a count in every preceding stimulus frequency-time bin (up to 250 ms) that marked a pip onset; thus, excitatory frequency-latency combinations produced above-average counts, while inhibitory combinations produced below-average counts. Counts were then converted to firing rates in units of spikes /s /pip, and the STRF was smoothed with a 5x5-bin uniform filter. The best frequency (BF) was that which evoked the maximum instantaneous firing rate, i.e., the maximum entry in the STRF matrix.

**First- and Second-Order PW Predictions of Measured tMTFs**

First- and second-order PW models were used to predict neuronal tMTFs, which were also derived directly from responses to periodic click trains with click repetition rates ranging logarithmically from 2 to 64 Hz (four rates per octave for a total of 21 rates). Each train began with a click at time zero and was 1 s in duration (so that the 2 Hz train consisted of 3 clicks), and was followed by 2 s of silence. Each train was repeated 10 times in a random se-
sequence (total ensemble duration was 10.5 minutes). Peak click intensity was 65 dB pe SPL, the same as that used in kernel estimation. The tMTF was defined as the average number of synchronized spikes evoked per click in the steady-state, as a function of the click repetition rate. We found that steady-state was always reached within 300 ms of a train’s onset; thus, all earlier clicks and spikes were excluded from the calculation. Synchronized firing was computed by Fourier transformation of the averaged period histograms, which were always constructed using 16 bins per period, so that the relative timing accuracy of the spikes was constant across click repetition rate (Eggermont, 2002). The significance of the synchronized response was assessed using the Rayleigh test of uniformity on its vector strength, a phase-locking metric (Fisher, 1993).

A first-order predicted response to a given periodic click train was obtained by summing the zeroth- and first-order functionals (Equations 3–4). Once computations reached steady-state, the predicted period histogram could be taken as the predicted response to any single click in a train, and the synchronized firing rate was computed using the same procedure outlined above for measured period histograms. Note that an equivalent first-order prediction of the tMTF can be computed simply by taking the Fourier transformation of the first-order kernel, underscoring the elegance of the convolution theorem. Likewise, a second-order predicted response was obtained by summing the zeroth-, first- and second-order functionals (Equations 3–5), and proceeding as above. Note that in the second-order case, transforming to the frequency domain would do little to simplify the computation (see, e.g., Bendat, 1998).

For both first- and second-order models, the percent predictability of the measured tMTF at a given click repetition rate was defined as 100 minus the percent error:

\[
\text{predictability} = 100 - \left( \frac{\text{Firing}_{\text{predicted}} - \text{Firing}_{\text{measured}}}{\max\{\text{Firing}_{\text{predicted}}, \text{Firing}_{\text{measured}}\}} \right) \times 100 \quad (10)
\]
The percent predictability will be reported at just the best click repetition rate or modulation frequency (BMF) of the measured tMTF, as well as averaged across those modulation frequencies at which the measured synchronized response was significant at p<0.05.

Case Inclusion Criteria

From all available MSU recordings, a smaller sample was chosen for analysis based solely on the criteria of response strength and stationarity. Stationarity was assessed using the Mann-Kendall test (Mann, 1945; Kendall, 1975), which detects monotonic trends in the inter-spike interval over time; recordings with a trend significant at p<0.05 were excluded as nonstationary. Minimum firing rate requirements were as follows. In response to both the Poisson-clicks and Poisson-pips stimuli, the peak firing rate of the (smoothed) first-order kernel had to be > 4 SD above the baseline rate, which was obtained by averaging spike activity over the first 10 ms post stimulus-onset (shorter than the minimum response transit time to AI). For periodic click trains, the synchronized response at the BMF had to be > 1 spike per click, and its vector strength significant at p<0.05.

Results

Proof of Concept: First- and Second-Order Poisson-Wiener Predictions of the Responses of Simulated Nonlinear Neurons

The firing properties of simulated AI neurons were specified as follows. The probability of a spike in each time bin was initially set to give a mean Poisson-distributed spontaneous rate of 20 spikes /s. In a simulated linear neuron, a stimulus click produced an average excitation (above the spontaneous rate) of 100 spikes /s over post-onset latencies of 20 to 40 ms, fol-
allowed by an average suppression (below the spontaneous rate) of 20 spikes /s over 40 to 100 ms. In a simulated second-order nonlinear neuron, the above impulse response was modulated by a simple constant function of the interval to every preceding impulse within the 200 ms memory of the system. Note that all of the above values are arbitrary, but representative of extracellular MSU recordings in AI. In the example presented as Figure 1, the second-order constant was set to ½; (values less than unity define a compressive nonlinearity and those greater than unity an expansive nonlinearity). Thus, the simulated nonlinear neuron’s average response to a click was ½ that of a linear neuron’s if the click was preceded by one other within the 200 ms memory, ¼ that of a linear neuron’s if the click was preceded by two others within memory, etc. The average Poisson-distributed click stimulation rate ($\lambda$ in Equations 3–8) was 4 /s, and each simulation was run over 30 minutes ($1.8 \times 10^6$ bins).

The top panel of Figure 1A shows the first-order kernel of the simulated linear neuron specified above; in this linear case, the first-order kernel is equivalent to the mean spike rate-corrected impulse response. The next panel shows the neuron’s second-order kernel, smoothed by a 21x21-bin uniform filter. The second-order kernel represents the average facilitation (by the red end of the color spectrum) or depression (by the blue end) of the impulse response at a latency of $\tau$ ms (horizontal axis), when the impulse is preceded by another impulse by $\theta$ ms (vertical axis), regardless of any intervening impulses. Even with 30 minutes of simulation, there is considerable noise in the second-order kernel estimate, which is related to the level of spike activity, but any substantial horizontal or vertical strip has a mean value of approximately zero (as required for a linear system). Note, however, that several artifacts (shown by white arrows) arise in our computation of the second-order kernel. These artifacts are a consequence of the discretization of Equation 8 and become smaller at shorter bin-widths. At our 1 ms resolution, the resulting second-order prediction errors were found to be less than 3%.
The third panel of Figure 1A shows the first-order predicted response to a four-click stimulus, drawn below; the prediction is, of course, a close match to the linear simulation specifications.

The top panel of Figure 1B shows the computed first-order kernel of a simulated second-order nonlinear neuron. Note that both excitation and suppression are smaller than in the linear case, due to the compressive nonlinearity of factor $\frac{1}{2}$. The neuron’s second-order kernel (next panel) shows an average depression of ~30 spikes /s /click during the 20 ms duration of excitation and extending over the 200 ms memory, and an average facilitation of ~6 spikes /s /click during the 60 ms duration of suppression (in this case, “facilitation” is better understood as “reduction of inhibition”). Note that 30:6 is also the ratio of excitation to suppression (100:20); note also that adding 30 spikes /s /click to the excitatory part of the nonlinear neuron’s first-order kernel, and subtracting 6 spikes /s /click from the suppressive part gives back the impulse response of the linear neuron. The first-order predicted response of the nonlinear neuron to the four-click stimulus is shown in the third panel of Figure 1B. Note that the first-order model underpredicts the responses to the first and fourth “linear” clicks, which are not affected by preceding clicks in memory, and overpredicts the responses to the second and third “nonlinear” clicks, which are affected by one and two preceding clicks, respectively. In contrast, the second-order prediction (bottom panel) is again a close match to the nonlinear simulation specifications, which require average excitatory responses of 70 (i.e., $20 + \frac{1}{2}[100]$) and 45 (i.e., $20 + \frac{1}{2}[2][100]$) spikes /s for the second and third clicks, respectively.

We carried out additional simulations to determine whether the expected second-order predictions would also be obtained for variations of several key, non-arbitrary parameters, namely the strength of the nonlinearity and the average Poisson-click stimulation rate. We
found that both parameters could only be increased within limits. For increased nonlinearity strength, this was a consequence of the increase in statistical error in the second-order kernel estimate due to the decrease in the number of spikes. For increased stimulation rate, it was a consequence of the increase in systematic error due to the increased deviation of the test-input distribution from the ideal Poisson (resulting from the 1 ms “dead time” imposed by discretization). As was the case for the first-order prediction of the response of a compressively nonlinear neuron, the systematic error due to the dead time was an underprediction of responses to “linear” clicks and an overprediction of responses to “nonlinear” clicks. It should further be noted that for systems of higher than second-order, a second-order Poisson-Wiener prediction will not be independent of the Poisson-click rate. This is a separate issue from that of the dead time and is a consequence of the decreased number of long inter-click intervals (as the click rate increases) with which the system is probed. This is not a problem for second-order systems where again, interactions occur only between pairs of clicks, regardless of any clicks in between.

First- and Second-Order Poisson-Wiener Kernels and Predictions: Examples from the Experimental Data

Extracellular recordings were made from AI in seven ketamine-anesthetized cats. The results reported here are based on 58 sets of multiple single-unit (MSU) recordings, which are presumed to come from large pyramidal neurons in layers III/IV of AI. Cases were selected from a larger pool based on the stationarity and the strength of responses to the Poisson-distributed and periodic click train stimuli (see Methods).

Figure 2 presents results from an example MSU recording (#352-2-3). Figure 2A shows its frequency tuning curve (FTC). The average response to a given tone pip frequency and inten-
The first-order kernel for the same example MSU, derived from responses to the Poisson-clicks test-stimulus, is given in the top panel of Figure 2C. As was the case for the vast majority of our recordings, the first-order kernel is characterized by a brief, strong excitation, which is followed by a longer, hard suppression (little or no spiking). In this example, there is also a weak rebound that peaks at ~190 ms. For comparison, kernels were also obtained from responses to the Poisson-pips test-stimulus (see Methods), and the Poisson-pips first-order kernel is shown in the top panel of Figure 2D. Note the qualitative similarity between the click- and pip-derived first-order kernels, albeit that the rebound activity peaks earlier with the pips (~150 ms). Note also that the Poisson-pips first-order kernel is computationally equivalent to the mean-rate-corrected time marginal of the STRF (for a proof, see Rieke et al., 1997), as can be appreciated by comparing the excitation-suppression-rebound time course of the STRF to that of the first-order kernel.

Second-order kernels derived from responses to the Poisson-clicks and Poisson-pips test-inputs are presented in the bottom panels of Figures 2C and 2D, respectively. Note, again, their qualitative similarity, except that the Poisson-clicks kernel is empty for $\theta \leq 20$ ms due to the dead time imposed between successive clicks (see Methods). The characteristic feature of these and in fact of all of our sampled second-order kernels was a marked (nonlinear) depres-
sion of the (linear) excitatory response. The depression was generally strongest at the shortest interstimulus intervals (again, represented by the blue end of the color spectrum), but could persist up to the longest intervals investigated (500 ms). This depression during periods of excitation was always coupled to a (nonlinear) facilitation (red colors) during periods of (linear) suppression. Put more concisely, all second-order kernels exhibited strong compressive nonlinearities, which acted to reduce the magnitude of both the excitatory and suppressive components of the first-order response. Expansive nonlinearities (i.e., the facilitation of excitation) were not observed. Our finding of only compressive nonlinearities in AI of anesthetized cats is consistent with the results of intracellular studies on large pyramidal neurons in various regions of rat sensory-motor cortex (Thomson et al., 1993; Thomson, 1997; Wehr and Zador, 2005). In contrast, more diverse nonlinear dynamics, including expansive nonlinearities, were observed in the auditory midbrain of the grassfrog (Epping and Eggermont, 1986), and in non-pyramidal neurons of cerebral cortex (e.g., Zucker and Regehr, 2002).

To evaluate the predictive power of first- and second-order PW models, we computed model responses to sets of periodic click trains with repetition rates of 2 to 64 Hz, and compared them to measured responses to the same stimuli (see Methods). We mention parenthetically that kernels estimated with the Poisson-pips test-input, while useful in many respects, cannot be used to generate quantitative predictions because different-frequency pips do not, of course, constitute equally effective impulse stimuli. Figure 2E shows, again for the same example MSU, the raw PSTH responses to the periodic click train stimuli, while the tMTF derived from these data is illustrated with the black curve in Figure 2F. The click repetition rate that elicits the most synchronized spikes /click, also referred to as the unit’s best modulation frequency (BMF) is 5.7 Hz (indicated by the red arrow in Figure 2E). Note that at low click repetition rates, this sharply-tuned bandpass unit exhibits strong periodic ringing at a fre-
frequency that closely matches the unit’s BMF. In Figure 2F, first- and second-order PW predictions of the tMTF are illustrated with purple and blue curves, respectively. We compared measured and predicted BMFs, as well as modulation frequency bandwidths at half-peak value. We also calculated the percent predictability of the tMTF, both at the measured BMF, and averaged across those modulation frequencies at which the synchronized response was significant (see Methods). In this unit, measured and first- and second-order predicted BMFs were all 5.7 Hz, and bandwidths were 1.0, 3.0 and 1.5 octaves, respectively. First-order predictability of the tMTF at the measured BMF was 70% and averaged first-order predictability was 78%. For the second-order model, these prediction values improved to 86% and 84%, respectively.

Figure 3 shows results obtained from three additional MSUs (A–C); STRFs are in the top row, first- and second-order PW kernels in the second and third rows, PSTH responses to periodic click trains in the fourth row, and measured and predicted tMTFs in the bottom row. Notice again that the (click-derived) first-order kernel closely describes the time course of the (pip-derived) STRF around the BF. Measured tMTFs were classified by eye as lowpass in 62% (36/58) of MSUs (e.g., Figures 3A, 3C) and as bandpass in the remaining 38% (22/58; e.g., Figures 2, 3B). Almost all bandpass units showed strong periodic rebound activity in PSTH responses at low click repetition rates (and in the first-order kernel), at frequencies that corresponded with the units’ BMFs. In each example unit, whether lowpass or bandpass, the second-order PW fit to the measured tMTF represents a substantial improvement over the first-order fit. One notices, however, that even the second-order models overpredict the tMTFs above ~16 Hz, implying that higher than second-order nonlinearities are present, and influence measured responses particularly at high stimulation rates. Parenthetically, we note that one-sixth (6/36) of lowpass tMTFs showed a pronounced notch at click repetition rates of
~3.4 Hz (e.g., Figure 3A), which could not be predicted with the PW models. The lack of spikes sometimes observed at this particular click rate is paralleled by the destructive interference seen in click-evoked local field potentials (LFPs). Thus, the negative, spike-generating phase of a given LFP is cancelled by the positive-going phase of the previous LFP, seemingly reducing the generation of spikes (Eggermont and Smith, 1995).

*First- and Second-Order Poisson-Wiener Kernels and Predictions: Averaged Results for Lowpass and Bandpass MSUs*

The top row of Figure 4 shows averaged tMTFs from the 22 bandpass (A) and 30 lowpass (B) MSUs (the six “lowpass with notch” MSUs were excluded, on this plot only); measured, and first- and second-order predicted tMTFs are color-coded as before. Prior to averaging, each case was normalized on the maximum synchronized firing rate of its measured and predicted data. The dashed black curves show ±1 SD about the mean measured tMTF. For the bandpass units, the average second-order prediction happened to be nearly exact up to modulation frequencies of ~14 Hz. For the lowpass units, the improvement over the first-order model was still substantial, though the average second-order prediction retained some of the bandpass tendencies of the first-order prediction, and was poorer than the second-order fit to bandpass tMTFs. The poorer fit is at least in part a consequence of the lower response SNRs, i.e., a lower ratio of stimulated to spontaneous activity, observed in our sample of lowpass compared to bandpass units (for reasons unknown). The mean peak excitation of the first-order kernels of lowpass units was 58 ± 18 (SD) spikes/s /click, and for bandpass units it was 78 ± 20 spikes/s /click (p<0.01; Student’s t-test). Figure 5 plots average second-order tMTF predictability versus response SNR, for the sample of lowpass (small, filled circles) and bandpass MSUs (large, open circles). SNR was taken as the ratio of the peak of first-order kernel excitation to the value of the zeroth-order kernel (i.e., the overall mean firing rate).
Significant correlations between predictability and response SNR were observed for both lowpass (solid line; $r^2=0.30; p<10^{-3}$) and bandpass units (dashed line; $r^2=0.24; p=0.02$). Mean ($\pm$ SD) SNRs were $18 \pm 9.0$ for lowpass units and $25 \pm 11$ for bandpass units ($p=0.05$; Student’s t-test).

A careful case-by-case inspection indicated that neither the general features of first- and second-order kernels, nor tMTF predictability, varied systematically with MSU frequency-tuning. Both did, however, vary with the type of measured tMTF. As already mentioned, first-order kernels of bandpass units typically showed some post-suppression rebound activity, whereas those of lowpass units did not. This is apparent from the averaged first-order kernels, shown in the second row of Figure 4 (note that individual cases were normalized on their maximum excitation before averaging), although the rebound is small as a result of the smearing caused by averaging (rebounds occurred at different latencies in different units). Also notable is that the decay of second-order depression with increasing interstimulus lag was typically nonmonotonic for bandpass MSUs (e.g., Figures 2C, 3B), but rarely so for low-pass MSUs (e.g., Figures 3A, 3C). This is also apparent from the averaged second-order kernels, shown in the third row of Figure 4 (individual cases were normalized on their maximum facilitation before averaging). The bottom row of Figure 4 illustrates the strength of second-order depression as a function of the interstimulus lag; at each lag ($\theta$ domain), the depression was averaged over latencies ($\tau$ domain) of 20 to 40 ms. The time course of the depression in lowpass units is reasonably well-fit by an exponential decay (thin curve) with a mean time constant of $93 \pm 18$ (SD) ms. The bandpass units are more roughly fit by an exponential decay with a time constant of $138 \pm 22$ ms, but there appears a significant secondary peak in the depression at an average lag of about 175 ms ($p<0.01$; Student’s t-test between the local minimum in depression at ~100 ms and the secondary peak at ~175 ms). No significant secondary
peak was observed in the lowpass group. Note that if the exponential fit is performed to only the first 100 ms of the data (i.e., up to the inflection point in the bandpass group), the mean time constants are a shorter 64 ± 15 ms for lowpass MSUs and 103 ± 21 ms for bandpass MSUs. Interestingly, as shown in Figure 6, there was a significant correlation between the BMF of a bandpass unit and the lag to the depression local minimum ($r^2=0.45$; $p<0.001$). Finally, in both the lowpass and bandpass groups, the depression strength remained above zero over the entire 500 ms period, but did so significantly for ~200 ms for lowpass MSUs and ~250 ms for bandpass MSUs.

The statistics (mean ± SD) of several parameters of measured and first- and second-order-predicted tMTFs, for the 22 bandpass and 36 lowpass MSUs, are presented in Table 1. These data further underscore the superior predictive power of the nonlinear compared to linear models. The average measured BMFs in the bandpass and lowpass groups were 6.6 ± 1.4 Hz and 3.1 ± 1.5 Hz, respectively. First-order models overpredicted the BMF of both groups, more so for lowpass units. On average, second-order predictions were not different from measured BMF values for bandpass units, but remained significantly higher ($p<10^{-4}$) than measured values for lowpass units. (All pair-wise comparisons were carried out using ANOVA and the post-hoc Tukey test.) Identical trends were obtained for predictions of tMTF bandwidths (see Table 1). The average predictability of the amplitude of the synchronized response at the measured BMF improved dramatically from first- to second-order models: from 58% to 76% for bandpass units and from 46% to 70% for lowpass units. Note that the improvement was actually greater in the lowpass group, though the second-order prediction was better in the bandpass group. Likewise, the average predictability of the entire tMTF improved from 69% to 78% (first- to second-order) for bandpass units and from 68% to 75% for lowpass units. The distributions of tMTF predictions from first- and second-order models are
presented in histogram form in Figure 7. Nonlinear predictions could be described as very good (≥ 80%) in almost half of our sample, but linear predictions only rarely so.

**Discussion**

We have demonstrated for the first time that an extension of the Wiener-Volterra theory to a Poisson-distributed impulse train input and spike output can be usefully applied to characterize the nonlinear response dynamics of neurons in primary auditory cortex (AI). There have been quite a few previous attempts to characterize the more peripheral stations of the auditory pathway using the second-order Wiener-Volterra approach. These include studies of the auditory nerve (Johnson, 1980; van Dijk et al., 1994; 1997a, b; Yamada and Lewis, 1999; Lewis et al., 2002; Recio-Spinoso et al., 2005; Temchin et al., 2005; Young and Calhoun, 2005), cochlear nucleus (Wickesberg et al., 1984; Recio-Spinoso and van Dijk, 2006), midbrain (Epping and Eggermont, 1986) and thalamus (Yeshurun et al., 1989). Of the above-cited studies, only those by Temchin et al. (2005) and Young and Calhoun (2005) investigated and reported significantly improved response predictions with second-order compared to first-order models. In the more distant past, one of the chief obstacles to a wider application of the Wiener-Volterra approach to sensory systems analysis had been its high computational demand.

Here we reported that second-order Poisson-Wiener (PW) models carry significantly more predictive power than do linear models in AI of the anesthetized cat. An even stronger finding is that second-order PW models provide very good fits to measured tMTFs in a majority of AI neurons, suggesting that the nonlinearities are predominantly of second-order. In our extracellular recordings, which presumably originated predominantly from large layer III and
IV pyramidal neurons, the observed nonlinearities were always compressive and generally decreased in strength over a memory of several hundreds of ms.

A well-known compressive nonlinear phenomenon with a memory of several hundreds of ms in the cerebral cortex, is often termed “forward suppression” (i.e., Calford and Semple, 1995; Brosch and Schreiner, 1997; Eggermont, 1999). There are several candidate mechanisms for forward suppression; chief among them are a GABA-mediated feed-forward inhibition (Wehr and Zador, 2005), and a Ca\(^{2+}\)-mediated short-term depression of synaptic transmission (Zucker and Regehr, 2002). Wehr and Zador (2005) demonstrated, in rat AI, that GABAergic inhibition could contribute to forward suppression over the first 50 to 100 ms or so, but that suppression could persist for up to 500 ms or more (as was confirmed here). Whether GABAergic inhibition is accompanied in AI by the type of thalamocortical synaptic depression which occurs in the visual (Freeman et al., 2002) and somatosensory systems (Chung et al., 2002) remains to be shown.

The first-order kernels of the majority of our bandpass units exhibited a periodic ringing whose frequency corresponded with the best modulation frequency (BMF) of the unit. In addition, second-order kernels showed distinct secondary peak in depression strength at inter-stimulus lags that also correlated with the BMF, i.e., the unit preferred modulation rates that coincided with a minimum in second-order depression. The mechanism responsible for this secondary, longer-latency nonlinear effect, which clearly contributes to the establishment of modulation tuning in AI, is not entirely clear. One possibility is autonomous periodic bursting in the spindle frequency range that is induced by ketamine anesthesia (Eggermont, 1992; Kenmochi and Eggermont, 1997). However, when we computed the power spectra of cortical local field potentials (LFPs; recorded from the 2 to 40 Hz signal on our electrodes) in the
absence of sound, and compared them with tMTFs measured on the same electrodes, we did not find a correlation between peaks in the LFP spectrum and peaks in the tMTF. In addition, on several occasions, we measured lowpass and bandpass tMTFs simultaneously on adjacent electrodes.

Using a phenomenological synaptic model with two time constants, one for the build-up of and another for the recovery from nonlinear depression, Eggermont (1999) obtained good fits to lowpass tMTFs measured in AI of the anesthetized cat. Good fits to bandpass tMTFs, however, necessitated the introduction of a third time constant for nonlinear facilitation, which is not supported by recordings from AI pyramidal neurons. The data presented here suggest that the model be amended to include a two-process depression. Other recent parametric models of short-term plasticity in neocortex also incorporate multiple depressive-facilitative mechanisms with distinct time constants (e.g., Markram et al., 1998; Song et al., 2008).

Given the substantial improvement in nonlinear over linear fits to both bandpass and lowpass tMTFs, it should be interesting to apply the Poisson-Wiener approach to the prediction of neural responses to stimuli outside the class used in kernel estimation, e.g., to conspecific vocalizations. A major problem with linear models is that they perform poorly in predicting responses to stimuli with different statistical properties than the characterizing stimulus class, even if their predictions to stimuli drawn from characterizing class are reasonably good (e.g., Kowalski et al., 1996; Sen et al., 2001; Woolley et al., 2006). The extent to which this limitation also applies to the second-order PW model remains to be addressed in future work.
**Acknowledgements**

Arnaud Noreña and Boris Gourévich assisted with the data collection. This work was supported by the Alberta Heritage Foundation for Medical Research, the National Sciences and Engineering Research Council of Canada, and the Campbell McLaurin Chair for Hearing Deficiencies.
References


Figure and Table Captions

Figure 1. First- and second-order PW kernels and response predictions for a simulated linear and a second-order compressively nonlinear neuron.

(A) First- and second-order PW kernels (top panels) for a simulated linear neuron (see the text for simulation specifications). In the linear case, the first-order kernel is equivalent to the impulse response. The second-order kernel illustrates the nonlinear depression and facilitation (by the blue and red ends of the color spectrum, respectively) of the impulse response due to interactions between pairs of impulses separated by time interval θ. Even the heavily smoothed (21x21-bin uniform filter) second-order kernel is noisy, but has a mean value of zero over any substantial vertical or horizontal strip for any linear neuron. Several artifacts due to the discrete computation appear in all second-order kernels, and are illustrated by white arrows. The first-order predicted response (purple curve) to a four-click stimulus (bottom panel) is a very close match to simulation specifications.

(B) First- and second-order PW kernels (top panels) for a simulated second-order compressively nonlinear neuron (see the text for simulation specifications). In this case, the first-order kernel is smaller than the impulse response and so the first-order model (purple curve) underpredicts the response to single clicks (i.e., those with no preceding clicks in memory), but overpredicts the response to multiple clicks. The second-order kernel characterizes and corrects for the compressive nonlinearity, so that the second-order prediction (blue curve) is a very close match to simulation specifications.

Figure 2. An illustration of the analytical procedures using data from an example MSU.

(A) Tone frequency tuning curves as a function of SPL, plotted as smoothed raster plots over latencies of 0 to 100 ms (time scale shown in the upper-right corner).

(B) STRF computed from responses to the Poisson-pips stimulus (see text).
(C–D) First- and second-order kernels computed from responses to the Poisson-clicks (C) and Poisson-pips (D) test-inputs. Note that the second-order Poisson-clicks kernels are empty for interstimulus lags (θ) of less than 20 ms, due to the minimum dead time imposed between successive clicks.

(E) Raw PSTH responses to periodic click trains with click repetition rates of 2 to 64 Hz. PSTHs represent the averages of 10 trials and were computed in 10 ms bins.

(F) Measured tMTF (black curve), and first- (purple) and second-order (blue) PW predictions.

Figure 3. Results from three additional example MSUs (A–C).
STRFs (top row), first- and second-order PW kernels (second and third rows), PSTH responses to periodic click trains (fourth row), and first- and second-order PW predictions of measured tMTFs (bottom row).

Figure 4. Averaged measured and predicted tMTFs (top row), averaged first- and second-order PW kernels (second and third rows), and averaged time courses of the strength of depression (bottom row) for the sample of 22 bandpass (A) and 36 lowpass (B) MSUs.
For each plot, individual cases were normalized on their response maxima before averaging. Dashed lines about the mean measured tMTFs and first-order kernels show ± SD. In the bottom row, the strength of the second-order depression at each interstimulus lag (θ domain) was averaged over latencies (τ domain) of 20 to 40 ms. For the bandpass but not for the lowpass units, the depression exhibits a significant secondary peak at ~175 ms.

Figure 5. Scatter plot of the SNR of the first-order response (taken as the ratio of peak to mean excitation) and the second-order predictability of the tMTF. Lowpass units are
represented by small, filled circles and bandpass units by large, open circles. Significant correlations between response SNR and predictability were observed for both lowpass (solid line; $r^2=0.30$; $p<10^{-3}$) and bandpass units (dashed line; $r^2=0.24$; $p=0.02$).

**Figure 6.** Scatter plot of the BMF and the interstimulus lag at which the second-order kernel of a bandpass MSU shows a local minimum in the strength of depression.

Data from 20 of the 22 MSUs that exhibited a clear local minimum is included. The correlation between the BMF and the time lag at the local minimum in depression is highly significant ($r^2=0.45$; $p<0.001$; regression line drawn on plot).

**Figure 7.** Histograms of the distributions of first- and second-order predictions of measured tMTFs (white and gray bars, respectively) for bandpass (A) and lowpass (B) MSUs.

**Table 1.** Averaged values (± SD) for bandpass (BP) and lowpass (LP) MSUs of the measured and predicted BMFs (first row) and tMTF bandwidths (second row), and of percent predictabilities of the tMTF, computed at the just the BMF (third row) and across all significant modulation frequencies (fourth row).
<table>
<thead>
<tr>
<th></th>
<th>Measured tMTF</th>
<th>Poisson-clicks Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First-order</td>
</tr>
<tr>
<td>BMF (Hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>6.6 ± 1.4</td>
<td>7.6 ± 1.2</td>
</tr>
<tr>
<td>LP</td>
<td>3.1 ± 1.5</td>
<td>7.8 ± 2.9</td>
</tr>
<tr>
<td>Temporal Modulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth (octave)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>1.7 ± 0.5</td>
<td>2.9 ± 1.0</td>
</tr>
<tr>
<td>LP</td>
<td>1.9 ± 0.8</td>
<td>3.5 ± 0.8</td>
</tr>
<tr>
<td>Predictability at the BMF (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td>58 ± 17</td>
</tr>
<tr>
<td>LP</td>
<td></td>
<td>46 ± 20</td>
</tr>
<tr>
<td>Average tMTF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td>69 ± 10</td>
</tr>
<tr>
<td>LP</td>
<td></td>
<td>68 ± 8.9</td>
</tr>
</tbody>
</table>

Table 1