

Auditory Edge Detection: A Neural Model for Physiological and Psychoacoustical Responses to Amplitude Transients

ALON FISHBACH,¹ ISRAEL NELKEN,¹ AND YEHEZKEL YESHURUN²

¹*Department of Physiology, Hadassah Medical School and Center for Neural Computation, Hebrew University, Jerusalem 91120; and* ²*Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel*

Received 19 June 2000; accepted in final form 21 February 2001

Fishbach, Alon, Israel Nelken, and Yehezkel Yeshurun. Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. *J Neurophysiol* 85: 2303–2323, 2001. Primary segmentation of visual scenes is based on spatiotemporal edges that are presumably detected by neurons throughout the visual system. In contrast, the way in which the auditory system decomposes complex auditory scenes is substantially less clear. There is diverse physiological and psychophysical evidence for the sensitivity of the auditory system to amplitude transients, which can be considered as a partial analogue to visual spatiotemporal edges. However, there is currently no theoretical framework in which these phenomena can be associated or related to the perceptual task of auditory source segregation. We propose a neural model for an auditory temporal edge detector, whose underlying principles are similar to classical visual edge detector models. Our main result is that this model reproduces published physiological responses to amplitude transients collected at multiple levels of the auditory pathways using a variety of experimental procedures. Moreover, the model successfully predicts physiological responses to a new set of amplitude transients, collected in cat primary auditory cortex and medial geniculate body. Additionally, the model reproduces several published psychoacoustical responses to amplitude transients as well as the psychoacoustical data for amplitude edge detection reported here for the first time. These results support the hypothesis that the response of auditory neurons to amplitude transients is the correlate of psychoacoustical edge detection.

INTRODUCTION

The sensitivity of the auditory system to amplitude transients is well documented, both physiologically and psychoacoustically. Psychoacoustical studies have demonstrated the importance of the temporal structure of amplitude envelope to auditory perception in general (e.g., Drullman 1995; Drullman et al. 1994a,b; Shannon et al. 1995; Turner et al. 1994), and to the segregation process of complex auditory scenes in particular (Bregman et al. 1994a,b). These studies demonstrate that both the magnitude and duration of amplitude transients affect auditory perception. However, it is still unclear which physical parameters of the amplitude transients most affect auditory perception of the transient.

Animal studies have shown that temporal changes in amplitude envelope in general, and amplitude onset in particular,

generate strong neural responses throughout the auditory pathways (Eggermont 1993; Kitzes et al. 1978; Phillips 1988; Rees and Møller 1983; Schreiner and Langner 1988a; Suga 1971). Several studies of the dependence of neuronal responses on the shape of an onset ramp (Barth and Burkard 1993; Heil 1997a,b; Heil and Irvine 1996, 1997; Phillips 1988, 1998; Phillips and Burkard 1999; Phillips et al. 1995) have shown that neural response characteristics can neither be ascribed to a simple function of onset plateau level nor to onset duration per se. Rather, the dynamics of the onset, such as the rate or acceleration of peak pressure, shape the neural response. These phenomena are evident across multiple levels of the auditory pathways. Furthermore, they have been demonstrated using a variety of experimental procedures, such as single-cell recordings from the cat primary auditory cortex and posterior field (Heil 1997a,b; Heil and Irvine 1996, 1998b; Phillips 1988, 1998), inferior colliculus potential of the awake chinchilla (Phillips and Burkard 1999), and human brain stem-evoked response (Barth and Burkard 1993).

The dependence of neural responses on the dynamics of the amplitude envelope raises the possibility that these responses reflect the computation of temporal auditory edges. Following this assumption, we suggest a neural model for the detection of amplitude transients (auditory temporal edges), which is inspired by visual edge detector models. The model responses are compared to published physiological responses to amplitude transients, and its predictions regarding the responses to amplitude transients that have not been examined before are verified experimentally. In addition, we attempt to define the physical parameters of amplitude transient that affect human perception of amplitude discontinuity, in order to characterize the psychophysical properties of perceived auditory temporal edge.

Our results suggest that the same physical parameters may govern both physiological and psychophysical responses to amplitude transients. Moreover, we show that both physiological and psychoacoustical responses can be explained by our simple neural model for auditory temporal edge detection. These results suggest that the sensitivity of the auditory system to amplitude transients is a realization of auditory temporal edge calculation that may have a primary role in neural auditory processing.

Address for reprint requests: I. Nelken, Dept. of Physiology, Hebrew University—Hadassah Medical School, PO Box 12272, Jerusalem 91120, Israel (E-mail: israel@md.huji.ac.il).

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

METHODS

Neural model principles

In line with the auditory-visual edge detection analogy, we adapted a model of visual edge detection to the auditory modality. The fundamental principle of the operation of visual edge detector is the calculation of a local brightness gradient. This is accomplished by differentiating the brightness function along some spatial direction or directions, using a combination of inhibitory and excitatory connections. The spatial organization of these connections in terms of the retinal image induces a receptive field that might be functionally described as an edge detector. Although there are recent and more elaborated visual receptive fields models, the simplest edge detecting receptive field model (Marr 1982; Rodieck 1965), which has an on-center off-surround (or vice versa) response pattern, suffices for our purpose. This receptive field describes the responses of edge detector neurons that can be found mostly in sub-cortical visual centers. The spatial properties of an idealized receptive field can be approximated by the second derivative of a gaussian or a difference of two gaussians (DOG), one wider than the other.

To adapt such a mechanism to auditory temporal edge detection, we hypothesize the existence of a temporal delay dimension, analogous to

the visual spatial dimensions. The stimulus is progressively delayed along this delay dimension. Information related to the temporal dynamics of the amplitude envelope (e.g., its rate of change) can be made explicit by differentiating the stimulus along this dimension, as the visual brightness gradient is made explicit by differentiating the stimulus along a spatial dimension.

We construct the delay dimension by using the well-known temporal characteristics of a standard version of the integrate-and-fire model (I&F). Our I&F makes use of a kernel function in the form

$$K(x) = \frac{1}{\tau_m} x e^{-x/\tau_m} \quad x \geq 0 \tag{1}$$

The kernel function, when convolved with the neuron's presynaptic input, determines its postsynaptic potential (Gerstner 1999a). τ_m is the membrane time constant that may range from 3 to 25 ms (McCormick et al. 1985). Higher τ_m values induce greater delay in the neuron's response (Agmon-Snir and Segev 1993). Inducing a receptive field in the delay dimension can be done by connecting the neurons with increasing τ_m s to an edge detector neuron using inhibitory and excitatory connections with various efficacies that reflect the receptive field shape. Differentiation of the stimuli is

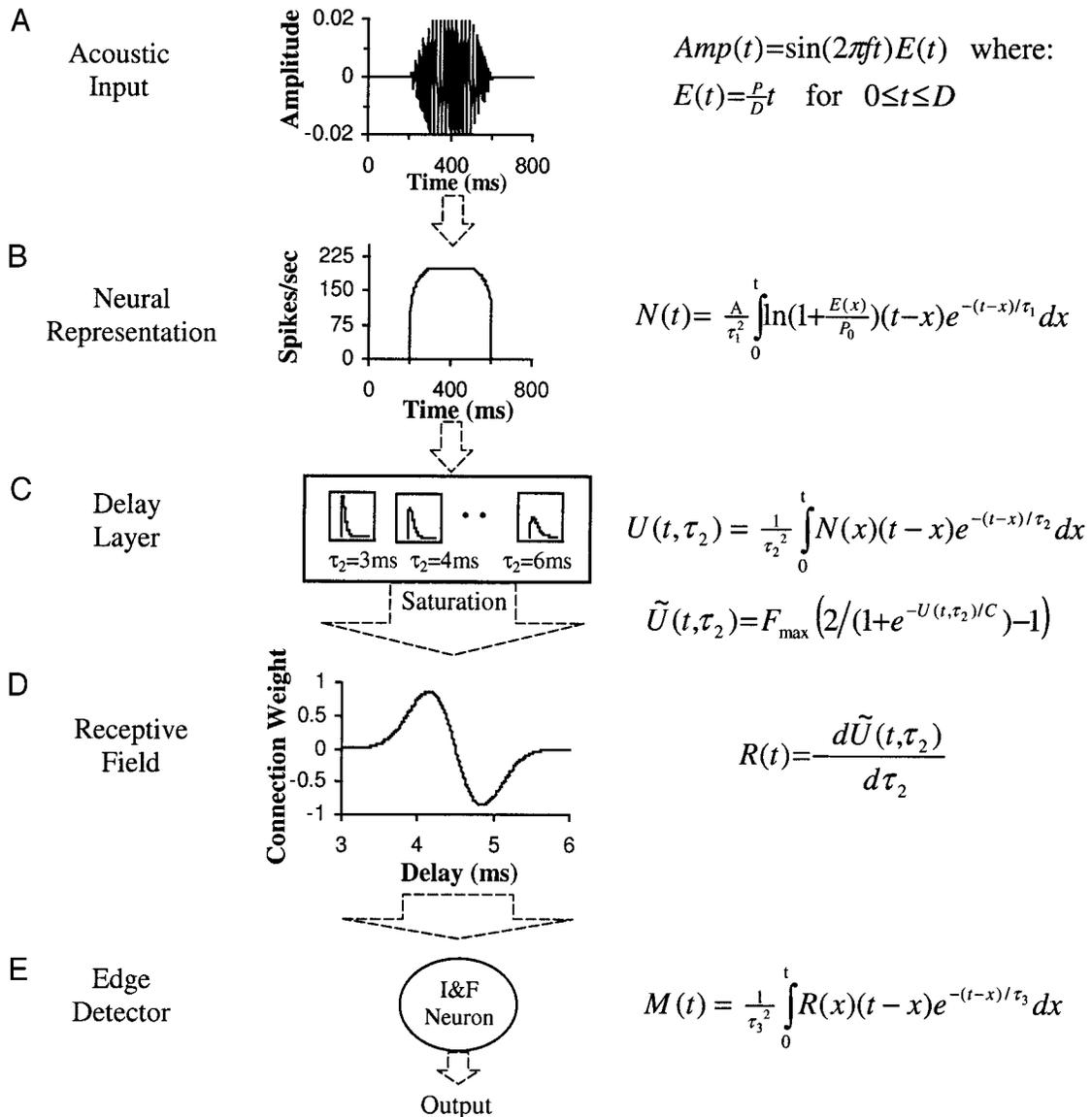


FIG. 1. Schematic diagram of the model. See detailed explanation in METHODS.

obtained by using a receptive field shape of a first-order derivative of a gaussian.

Figure 1 presents a schematic diagram of the model and the flow of data along the different model components. Each model component is annotated with an approximate expression for its operation on its input. These formulations will be used in the analysis of the model. Exact implementation details are given in APPENDIX A. The inputs tested consisted of tone bursts shaped with ON and OFF ramps of various shapes. An example of a tone burst with linear ramps is displayed in Fig. 1A.

NEURAL REPRESENTATION. The neural representation (Fig. 1B) is roughly the expected peripheral representation of sound by the inner hair cell DC potential. This representation is generated using a simple preprocessing that includes demodulation to extract the temporal envelope, non-linear compression and low-pass filtering. In our analysis we formulate the demodulation and the non-linear compression using the amplitude envelope of the input converted to dB SPL scale (the constants in Fig. 1B are set to $A = 20/\ln(10)$ and $P_0 = 2 \cdot 10^{-5}$ Pa). The form of the argument to the log transformation eases the analysis for near-zero values of t and has negligible effect for $t \gg 0$. The low-pass filtering is formulated by convolving the log-envelope with an alpha kernel function (Eq. 1) with a time constant of τ_1 , which is in the millisecond range (Hewitt and Meddis 1990; Smith 1988). This preprocessing stage can be replaced by a more realistic inner hair cell model (which produces simulation of auditory nerve firing probabilities) (Hewitt and Meddis 1990; as implemented by Slaney 1998) without any qualitative change in the response characteristics of the model.

DELAY LAYER. The preprocessed input is fed to the delay layer of the model, which consists of standard integrate & fire (I&F) neurons with ascending membrane time decay constants. Each unit $U(t, \tau_2)$ in the delay layer represents a population of neurons with identical characteristics. The population response is modeled as an analogue variable, by convolving the neuronal representation $N(t)$, with a kernel (Gerstner 1999b) whose time constant is τ_2 . I&F kernel functions and membrane time-constant values are shown for several units (Fig. 1C). The membrane potential of each neuron in the delay layer is then saturated using a sigmoidal function

$$S(x) = F_{\max} \{2/[1 + e^{-(x/C)}] - 1\} \quad (2)$$

where F_{\max} is the maximal instantaneous output firing rate (225 spikes/s) and C is a scaling factor, which determines the dynamic range of the transformation. In Fig. 3, the outputs of the delay layer neurons (including various amounts of saturation) are shown for stimuli, similar to the stimulus presented in Fig. 1.

RECEPTIVE FIELD. The delay layer neurons are connected to an edge detector neuron using inhibitory and excitatory connections with various efficacies (Fig. 1D) that reflect the receptive field shape, which is a first derivative of a gaussian. The output of the receptive field, $R(t)$, is shown in Fig. 3 for stimuli similar to the stimulus presented in Fig. 1 and is approximately a smoothed first derivative of the outputs of the delay neurons along the τ_2 dimension.

EDGE DETECTOR NEURON. The edge detector neuron (Fig. 1E) is a single I&F neuron with a membrane time constant τ_3 . The output of the edge detector neuron is also the output of the model. In the numerical implementation of the model, a noisy integration was used (Gerstner 1999a). For the analytical treatment presented here, the membrane potential of the edge detection neuron, $M(t)$, is modeled as a low-pass filter operating on the output of the receptive field operator, $R(t)$.

PARAMETERS OF THE MODEL. The responses of the model are adjusted to fit the response of a specific neuron by adjusting two parameters. The first parameter is C , the scaling factor of the delay layer saturation transformation (Eq. 2), and the second parameter is τ_3 , the membrane time constant of the edge detector neuron. In addition,

the threshold of the edge detector neuron was varied. However, the threshold was not manipulated independently; instead, its value was always set to best approximate the threshold of the neuron that was fitted. There are six additional fixed parameters of the model; three of them are parameters of the I&F model. These parameters and their values are listed in full in APPENDIX A. Their specific values have only minor or redundant effect on the responses of the model. For example, changing the value of τ_1 or the range of τ_2 that are used in the delay layer can be in large extent be compensated by adjusting the value of τ_3 .

Physiological methods

ANIMALS AND PREPARATION. Neurons have been recorded in primary auditory cortex (AI) and medial geniculate body (MGB) of two halothane-anesthetized adult cats. The methods have been described in details elsewhere (Nelken et al. 1999). In short, the cats were premedicated with xylazine (0.1 ml im), and anesthesia was induced by ketamine (30 mg/kg im). The radial vein, the femoral artery, and the trachea were cannulated. Blood pressure and CO_2 levels in the trachea were continuously monitored. The cat was respired with a mixture of $\text{O}_2/\text{N}_2\text{O}$ (30%/70%) and halothane (0.2–1.5%, as needed). Halothane level was set so that arterial blood pressure was kept around 100 mmHg on the average. Under these conditions, the cat usually could be respired without the use of muscle relaxants. In case muscle relaxants were required, the depth of anesthesia was evaluated by testing paw withdrawal reflexes before administering low levels (pancuronium bromide, 0.05–0.1 mg iv, typically once every 2–3 h). Lactated ringer was continuously given through the venous catheter (10 ml/h). Every 8–12 h a chemical analysis of arterial blood was performed. When the cat developed acidosis, bicarbonate was given (typically 5 ml iv, every 8 h).

AI was accessed using standard methods. To reach the MGB, electrodes were introduced at the appropriate stereotactic coordinates. Physiological characteristics of the neuronal activity were used to position the electrode at the ventral division of the MGB. The electrodes were stained with DiI, and the localization was verified after the experiments using histological reconstruction of the electrode tracks. The animal protocol was approved by the local animal care committee.

DATA ACQUISITION. Glass-coated tungsten electrodes (locally made) were used for recording neuronal activity. The activity from the electrodes was amplified (MCP8 Plus, Alpha-Omega), and spikes were detected on-line by a spike sorter (MSD, Alpha-Omega). The times of the spikes were recorded (ET1, TDT) and written into a file for off-line analysis.

ACOUSTIC STIMULATION. Stimuli were generated digitally converted to analog waveforms and attenuated using TDT equipment. All stimuli were tone bursts, 230 ms long including the symmetrical onset and offset ramps. Six types of onset/offset window shapes were used, $\cos^2(t)$, $\cos^4(t)$, t , t^2 , t^4 , and squared exponential. By denoting the plateau peak pressure in Pascal units with P , and the onset rise time in milliseconds with D , the peak pressure (in Pa) during the onset is given by

$$E_c(t) = P \left(\frac{t}{D}\right)^n \quad \text{for } 0 \leq t \leq D \quad n = \{1, 2, 4\} \quad (3)$$

for the t , t^2 , and t^4 windows, and is given by

$$E_c(t) = P \cos^n \left(\frac{t\pi}{2D} + \frac{\pi}{2}\right) \quad \text{for } 0 \leq t \leq D \quad n = \{2, 4\} \quad (4)$$

for the $\cos^2(t)$ and $\cos^4(t)$ windows. For the squared exponential window, the peak level (in dB instead of in Pa) is given by Eq. 3 with $n = 2$, except that P is given in dB. To accommodate the peak pressure close to 0 Pa (at the beginning of the onset and the end of the

offset), where the dB scale is singular, a short linear ramp was used up to peak sound levels of about 0 dB SPL.

Onset window shapes were generated either using an electronic switch (SW2, TDT) or in the digital domain (for the squared exponential windows). The sound was presented to the animal through electrostatic earphones (Sokolich) whose frequency response varied by less than 10 dB in the frequency range used here. In situ calibration of the earphones was performed in each ear.

For the data presented here, neurons were presented with tone bursts at their best frequency. Tone levels were chosen from about 10 dB below neuronal threshold and up to about 100 dB SPL, in 10 dB steps. Tone rise times covered the range of 1.7–100 ms and were measured between 10 and 90% amplitude points when generated using the electronic switch, or between 0 and 100% amplitude points when generated in the digital domain. Data were taken in blocks, within which the window shape was kept constant, but the tone level varied randomly under the constraint that each level was presented 20 times. Stimuli were presented at a rate of 1/s. After a block was finished, another window shape (or a different rise time) was selected, and the process was repeated. In total, 19 neurons in AI and 9 neurons in MGB were tested with these stimuli. Of these, data from 11 neurons in AI and 4 neurons in MGB, whose responses were strong and stable during the recording session, were analyzed for this paper.

Psychoacoustical methods

The main goal of our psychoacoustical experiments was to test whether the perception of amplitude changes is determined by the gradient of the change, or by some other combination of its duration and magnitude. A secondary goal was to rule out the possibility that the sensitivity of the auditory system to amplitude changes is due to a spectral splatter that may be induced by the sudden amplitude change. In order to accomplish these goals we used a direct measure of the way in which the amplitude change is being perceived, rather than measuring amplitude change effect on higher perceptual tasks. This enabled us to isolate the perception of the amplitude transient from the context of more elaborate auditory phenomena such as auditory source segregation, in order to avoid high-level cognitive influences. Two sets of experiments were conducted; the first measured the discontinuity perception of ramped sinusoids (*experiment 1*), while the second measured the perception of ramped noise bursts (*experiment 2*).

PARTICIPANTS. All participants were normal hearing volunteer adults, who participated with full informed consent. Data for *experiment 1* were obtained from 10 participants. All except for one, who is one of the authors (YY), had no previous listening experience in psychoacoustical experiments. Data for *experiment 2* were obtained from five participants. None had participated in *experiment 1*, and none had previous listening experience in psychoacoustic experiments.

STIMULI. *Experiment 1* stimuli are pure tones with an amplitude envelope as illustrated in Fig. 2 (solid line). Onset and offset times are 150 ms, and both plateau amplitude periods are 1 s. The first plateau level (A_1), the amplitude ramp size (ΔA) and duration (ΔT), and the frequency of the tone were manipulated. The values used appear in Table 1. The set of stimuli is a full combination of the variable's values, thus forming a set of 224 unique stimuli, each of which was presented once. The stimuli were generated digitally and played over a Silicon-Graphics Indigo workstation at sampling rate of 16,000 Hz at 16-bit resolution.

The stimuli used in *experiment 2* were prepared by Olsen (1994). All stimuli were broadband noise bursts, 700 ms in duration, 0–22 kHz bandwidth, uniform random, digitally generated using a PC computer and signal processing software (Signal, Engineering Design). The amplitude envelope of the noise burst was shaped by

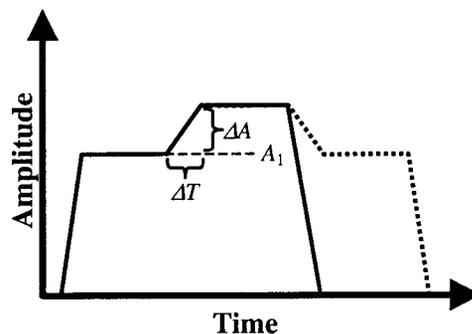


FIG. 2. Illustration of the amplitude envelopes of the stimuli that were used in *experiments 1* (solid line) and *2* (dashed line). In both experiments ΔA , ΔT , and A_1 were manipulated. Note that the time scales for the 2 experiments are different. The total duration of the tone stimuli used in *experiment 1* is $2,300 + \Delta T$ ms, while the duration of the noise bursts used in *experiment 2* is 700 ms.

multiplying the signal with a trapezoidal function, which is illustrated in Fig. 2 (dashed line) and contained 96-ms onset/offset time and 104 ms of plateau level before and after the pedestal. The values of the variables used in this experiment can be found in Table 2. The set of stimuli is a full combination of the variable's values, thus forming a set of 36 unique stimuli, each of which was presented 5 times. Stimulus levels for both experiments were calibrated using General-Audio 1562-Z audiometer calibration set.

PROCEDURE. An identical procedure was used in both experiments. The stimuli were presented binaurally through Yamaha HP-2 earphones to the participants who were seated in a soundproof room. The psychophysical task was to judge whether the transition between the two plateau amplitude levels was a continuous or discontinuous one. The participants were asked to indicate their choice for each of the stimuli using a two-alternative forced choice procedure. A random training subset of 40 trials was presented to the listeners, followed by the entire set presented in random order. The listeners were unaware of the fact that the first trials were training trials. Participants had unlimited time to respond after each trial and were presented with the next trial 2 s after their response.

RESULTS

Neural model: general observations

The model was capable of reproducing all the physiological characteristics of onset responses in AI neurons. In particular, the model was capable to produce the shortening of latencies with increase in tone level, and was capable of generating both monotonic and non-monotonic rate-level functions.

Figure 3 illustrates the way the model responds to amplitude transients and the effect of the delay layer's saturation on the timing and strength of the responses. The log-compressed envelopes of linearly shaped 30-dB SPL and 90-dB SPL tone bursts are shown in Fig. 3A. The response of the model components to these stimuli is considered in two different saturation conditions. A model with a highly saturated delay layer, which yields non-monotonic responses, is described in Fig. 3, B, D, and F, while a model with only weakly saturated delay layer is described in Fig. 3, C, E, and G. For clarity, we consider a simplified delay layer that consists of only two neurons with time constants of 3 and 6 ms. Figure 3, B and C, demonstrates the different delays of the stimulus envelope that are being induced by the two neurons. The outputs of the two neurons are subtracted by connecting them to the edge detector neuron with weights of

TABLE 1. *Values of variables used in experiment 1*

ΔT , ms	6	12	18	24	40	96	120
ΔA , dB	6	12	18	24			
f , Hz	250	400	650	950			
A_1 , dB SPL	60	73					

equal magnitude and opposite signs (Fig. 3, *D* and *E*). The prominent effect of the amount of saturation on the model responses emerges at this stage. For example, the total current (the integrated presynaptic input) that is being injected to the edge detector neuron in the highly saturated model (Fig. 3*D*) is higher in response to the 30-dB tone than to the 90-dB tone (155 vs. 89.7 in arbitrary units, respectively). In the weakly saturated model (Fig. 3*E*), the integrated presynaptic input is lower in response to the 30-dB tone than the 90-dB tone (81.9 vs. 246.8, respectively). The non-monotonicity of the highly saturated model is enhanced by the low-pass properties of the membrane potential of the edge detector neuron (Fig. 3*F*). The effect of the delay layer's saturation on the non-monotonicity of the model is being mathematically analyzed in APPENDIX B. Another effect of the saturation is decreasing the first-spike latency and shortening the period of neural activity. For the purpose of mathematical treatment, it can be reasonably assumed that a neuron starts to fire when its membrane potential hits a fixed threshold and that its spike count is proportional to the area enclosed by this threshold and the neuron's membrane potential (Fig. 3*G*).

Evaluation of the neural model: single-neuron data

We evaluated the adequacy of the model to match reported neural response to sound bursts by feeding the model with the amplitude envelope of the stimuli and comparing several aspects of the model output with those of the reported responses. The properties of the output examined were the first spike latency of the response, the response strength measured by the number of spikes that followed a stimulus, and the relationships between the two.

LATENCY. Heil and his co-workers (Heil 1997a; Heil and Irvine 1996) studied the latency of primary auditory cortex neurons (AI) as a function of the shape, amplitude, and duration of the rise time of a best frequency tone. Two kinds of onset envelope functions were used, linear and cosine-squared. The peak amplitude during a linear onset is described by a power function as described in Eq. 3 with $n = 1$. The peak amplitude during a cosine-squared onset is described by Eq. 4, with $n = 2$.

As was stated earlier, the main finding of Heil and his co-workers is that mean latency is not solely a function of one parameter of the onset envelope, but rather a function of the dynamics of the envelope. The latency of response appears to be a function of the rate of rise of the onset when a linear shaped onset is used, and a function of maximal acceleration of the envelope for cosine-squared onsets. Moreover, Heil proposed a functional expression for the relationships between the response latency and rate of rise (for linear onsets) or maximal

acceleration of peak pressure (for cosine-squared onsets). The function for the linear case is given by

$$L_l = L_{\min} + A_l * \left[\log \left(\frac{P}{D} \right) + S \right]^{-4} \quad (5)$$

where A_l is a global scaling factor, and L_{\min} and S are neuron specific parameters that determine the minimal latency of the neuron and its sensitivity to onset rate of rise, respectively.

The function for cosine squared onsets is given by

$$L_c = L_{\min} + A_c * \left[\log \left(\frac{\pi^2 P}{2D^2} \right) + S \right]^{-4} \quad (6)$$

Note that the term $\pi^2 P/2D^2$ stands for the maximal acceleration of the envelope, which occurs at the beginning of the onset. Heil fit global scaling factors A_l and A_c over the entire neural population that was recorded and set them to 1,277 and 12,719 ms, respectively.

We fitted the model parameters to match the responses of 13 AI neurons for which both latency and spike-count data are fully reported by Heil (1997a,b). For all of these neurons we found that the model reproduced the latency phenomena that were measured by Heil. The latency data for two of these neurons is shown in Fig. 4.

Figure 4, *A* and *B*, shows the experimental vs. simulated iso-rise-time curves of first-spike latency as a function of amplitude peak pressure of a cosine-squared onset. Figure 4, *C* and *D*, demonstrates that plotting both the experimental and simulated latency as a function of maximal acceleration of the cosine-squared onset brings the iso-rise-time curves to close congruence along a single curve that can be fitted by Eq. 6. Figure 4, *E* and *F*, shows the congruence of the iso-rise-time curves as a function of rate of linear rise onset.

Phillips (1998) and Heil and Irvine (1998b) reported the responses of single neurons in the cat primary auditory cortex and the posterior field to characteristic frequency (CF) tones with cosine-squared-shaped onsets. These data confirmed the initial observations of Heil and his co-workers in AI and extended them to a secondary cortical field. Figure 5, *A* and *C*, replots the first-spike latency of two neurons from the posterior field as reported by Phillips, and Fig. 5, *B* and *D*, plots the fit of the model to this data. Plotting Phillips' latency data as a function of maximal acceleration of the cosine-squared onset demonstrates again the close congruence of the latency data along a single curve, which can be fitted by Eq. 6.

FIXED-THRESHOLD MODEL DOES NOT FIT THE DATA. A possible explanation for the latency phenomena is that the neuron first spike occurs when the input stimuli level hits a fixed threshold (Kitzes et al. 1978; Phillips 1988; Suga 1971). Indeed, it is easy to show that such a simple model predicts a reciprocal relation between the first-spike latency of a neuron and the rate (P/D) of linear onsets and maximum acceleration ($\pi^2 P/2D^2$) of cosine-squared onsets. While these predictions roughly approximate the experimental results, the later systematically deviate from the predictions. On these grounds, Heil and Irvine (1996)

TABLE 2. *Values of variables used in experiment 2*

ΔT , ms	3	6	12	24	40	96
ΔA , dB	12	18				
A_1 , dB SPL	59	65	71			

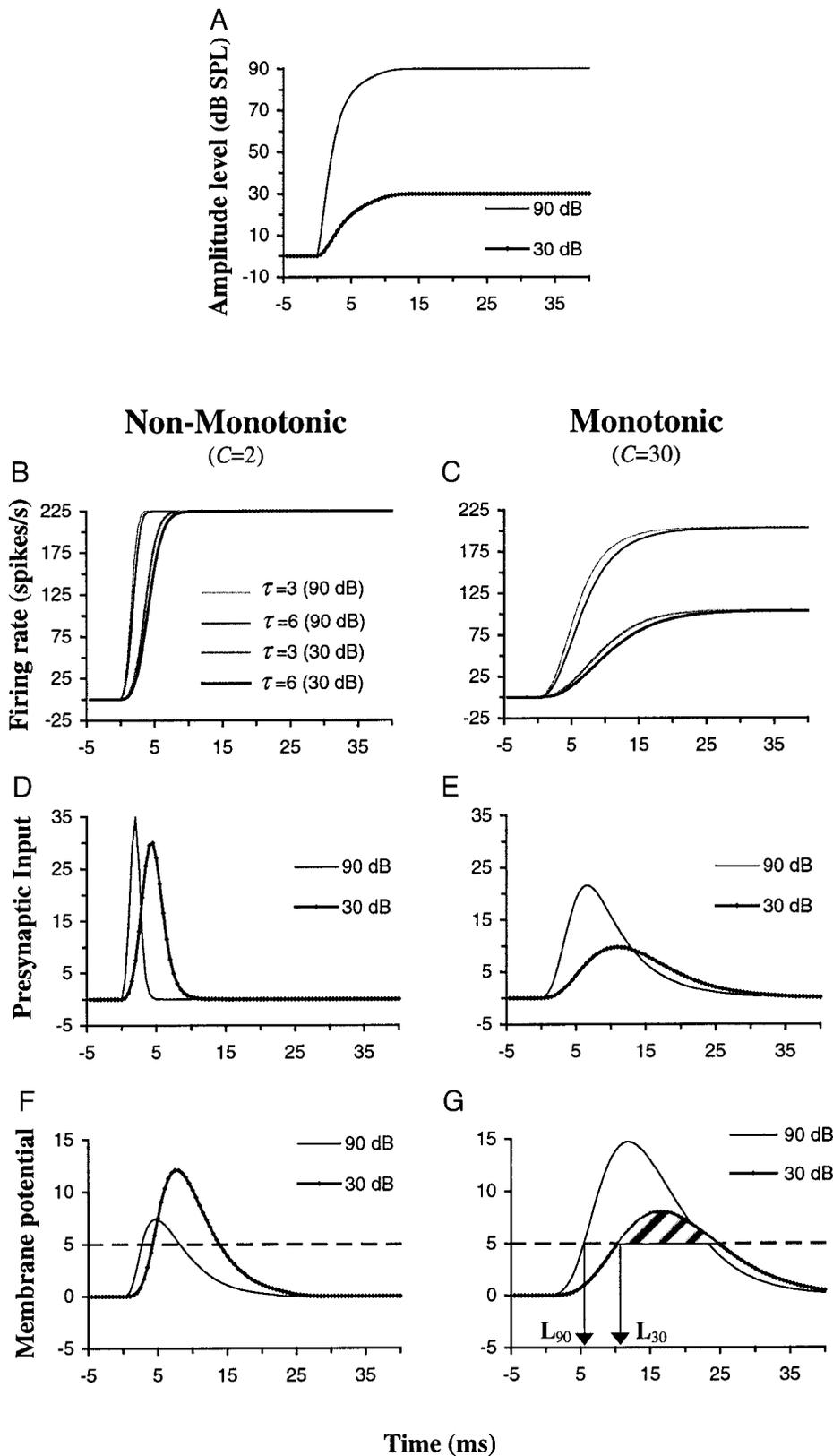


FIG. 3. The output of several of the model components as response to sound bursts of 2 amplitude levels (A). Two model settings are shown, the 1st includes highly saturated delay layer (B, D, and F), while the 2nd is only moderately saturated (C, E, and G). The figure demonstrates how the input is being progressively delayed along a simplified delay layer (B and C) and differentiated using the receptive field (D and E) that is formed by the connections from the delay layer neurons to the edge detector neuron (F and G). For mathematical analysis of the model we define the 1st-spike latency of the model as the time from stimulus onset to the 1st time the edge detector membrane potential hits a fixed threshold level (L_{30} and L_{90} in G). The spike count is assumed to be proportional to the area enclosed by the membrane potential and the threshold level (striped area in G).

argue against the simple threshold model. Their claims can be summarized by two main points that are illustrated in Fig. 6.

First, the threshold model predicts that the first-spike latency should be a linear function of rise time (see dashed lines in Fig. 6A). The experimental data of Heil and Irvine (1996; Heil

1997a) show a systematic deviation from this prediction. Notably, the relation between the latency and the rise time is compressive, which rules out the possibility that adaptive processes are the cause for this deviation.

The second argument relates to the slopes of the quasi-linear

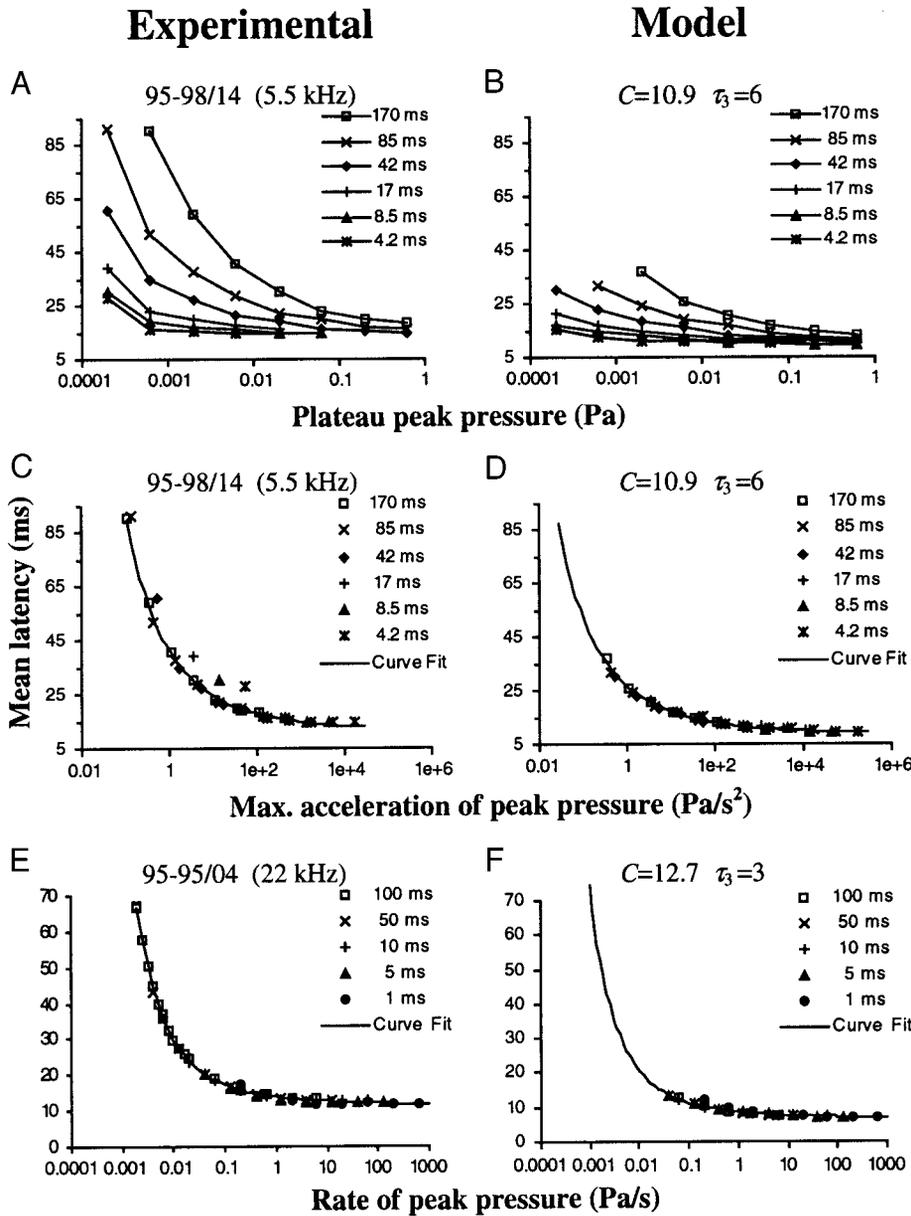


FIG. 4. Experimental (replotted from Heil 1997a) vs. model simulated data for 1st-spike latency of the onset response. *A* and *B*: the latency as a function of the amplitude level. *C* and *D*: the latency as a function of maximal acceleration of the cosine-squared onset and the curve fitted by Eq. 6. Our fit for the neuron in *C* yielded $S = 4.53$ and $L_{\min} = 10.87$ ms, and for the simulated data in *D* the best fit yielded $S = 5.12$ and $L_{\min} = 8.1$ ms. *E* and *F*: the latency as a function of the rate of linearly shaped onset and the curve fitted by Eq. 5. The fit for the neuron in *E* yielded $S = 4.9$ and $L_{\min} = 11.55$ ms and for the model (*F*) $S = 5.09$ and $L_{\min} = 6.7$ ms. Neuron identity, neuron characteristic frequency (CF), and the model parameters that were used are shown above each plot. The difference between experimental and simulated L_{\min} values reflects constant delays (acoustic, cochlear, and neural delays), which are not included in the model. Model S values are consistently somewhat higher than those estimated from the data, as explained in DISCUSSION.

iso-step-size curves, which should decrease, according to the fixed threshold model, as the inverse ratio of the step sizes. Heil and Irvine demonstrate that the slopes of the curves decrease by a factor that is smaller than expected. Similar deviation from the threshold model predictions have been observed in the first-spike latency of cortical neurons as response to cosine-squared onsets (see Heil 1998 for reanalysis of the data of Phillips 1998); in the response latency of inferior colliculus potential in unanesthetized chinchillas to cosine-squared onsets (Phillips and Burkard 1999) and in the response latency of evoked cortical potentials in humans as response to linear onsets (Onishi and Davis 1968). Since our model reproduces very accurately the reported latency phenomena, it also shows deviations from the predictions of the fixed threshold model (Fig. 6B).

MATHEMATICAL ANALYSIS OF LATENCY PHENOMENA. In our analysis we use the formulations given in Fig. 1, and we assume that the amplitude envelope of the input stimulus, $E(t)$,

can be approximated during the onset (for $t \leq D$) by a power function such as described in Eq. 3 for any $n > 0$. For simplicity sake we will restrict our analysis to $t \leq D$; this assumption is equivalent to the statement that the first spike occurred during the onset ramp (after taking into account constant latency components that are independent of the sound level).

As illustrated in Fig. 3G, we assume that the edge detector neuron starts firing when its membrane potential, $M(t)$, hits a fixed threshold level, T . Thus the time of the first spike, t^* , satisfies the condition: $M(t^*) = T$. Although t^* can be calculated numerically using the implicit functional form $M(t^*) = T$ (as it is actually done in the process of fitting the model free parameters to match the experimental data), we are unable to extract an explicit expression for t^* that can replace Heil's functional forms (Eqs. 5 and 6). However, the implicit functional form is useful in order to prove several characteristics of experimental and simulated latency phenomena, and to predict

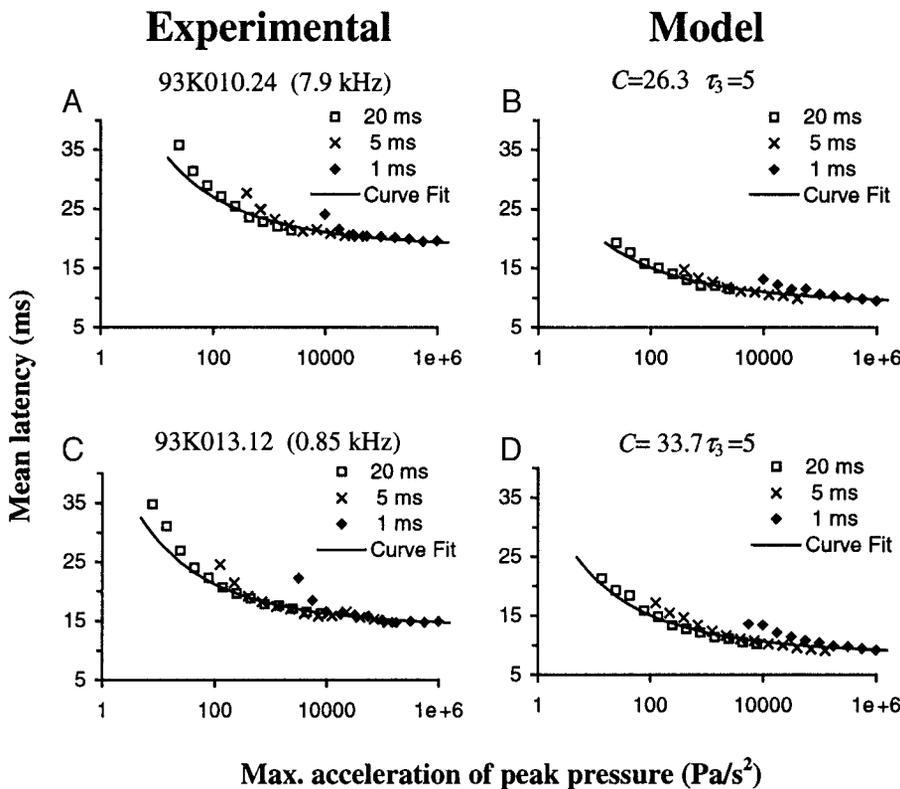


FIG. 5. Latency data from the cat posterior field (replotted from Phillips 1998) vs. model simulated data. The latency is plotted as a function of maximum acceleration of peak pressure and is fitted using Heil's functional form. Our estimated S value for neuron 93K010.24 (A) is 3.99, and the estimated value for the corresponding simulated latency (B) is 4.62. Estimated S for neuron 93K013.12 (C) is 4.42 and for the corresponding model setting (D) the estimation is 4.57.

latency behavior as response to stimuli that were not examined experimentally.

Since $M(t)$ includes only non-linear compression and linear time-invariant filtering of $E(t)$, it is clear that t^* , as a function of P and D , is being determined uniquely by the term P/D^n . This explains Heil's findings regarding the latency being a function of the rate of linear onsets ($n = 1$) while being a function of the maximum acceleration of cosine-squared onset ($n = 2$, up to a 1st-order approximation). In addition, this conclusion predicts that for a large family of functions that can be approximated by a power function, the first-spike latency for tone bursts that are shaped using these functions should be determined by the term P/D^n . Moreover, we predict that for exponential power functions, such that the envelope is a power function when P is given in dB units, t^* is determined by the term P/D^n , when P is given in dB units.

Note that the analysis in the previous paragraphs is limited to $t^* \leq D$ (1st spike generation occurring during the onset

ramp). For near-threshold levels of P , t^* may exceed D , which results in longer latencies than predicted. This presumably is the cause of the departures from the invariant relationship between first spike latency and P/D^n at low levels of P in both experimental and simulated data (e.g., Figs. 4C and 5).

Another phenomenon that can be explained by the implicit form of t^* is of Heil and Irvine (1996) regarding the deviations of the latency from the predictions of a fixed threshold model. In APPENDIX B we explore the dependence of t^* on the duration of the onset, D , and prove the compressive nature of $t^*(D)$ as evident in both experimental and simulated data (Fig. 6).

It should be noted that the latency of any fixed-threshold system, which includes only monotonic transformations and linear time-invariant filtering of $E(t)$, as a function of P , D , and n , is being uniquely determined by the term P/D^n . This observation can account for the latency phenomena of auditory nerve fibers, reported by Heil and Irvine (1997).

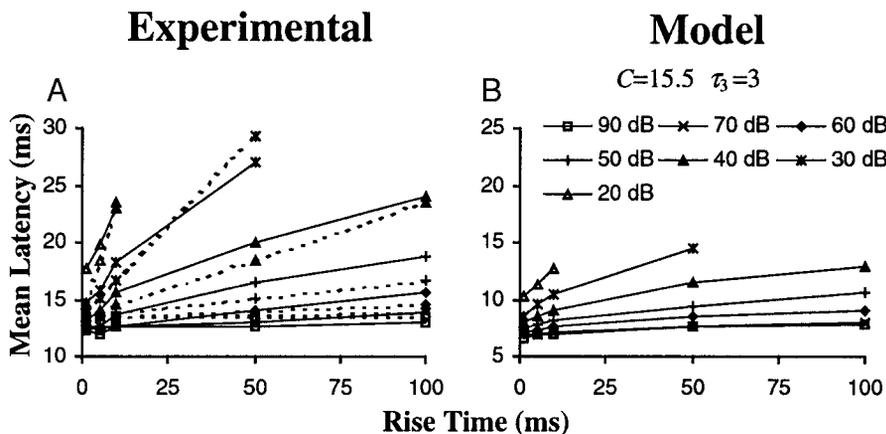


FIG. 6. A: mean 1st-spike latency of a neuron from cat primary auditory cortex (solid lines) as a function of rise time of a CF tone of 22 kHz (replotted from Heil and Irvine 1996). The dashed lines plot the best fit of the data according to a fixed-threshold model. Note that the latency is not a linear function of rise time and that the slope ratio of any two quasi-linear iso-step-size curves does not match the inverse ratio of the step sizes. The model reproduces these phenomena (B). Note that the latency axis is translated with respect to A for greater clarity.

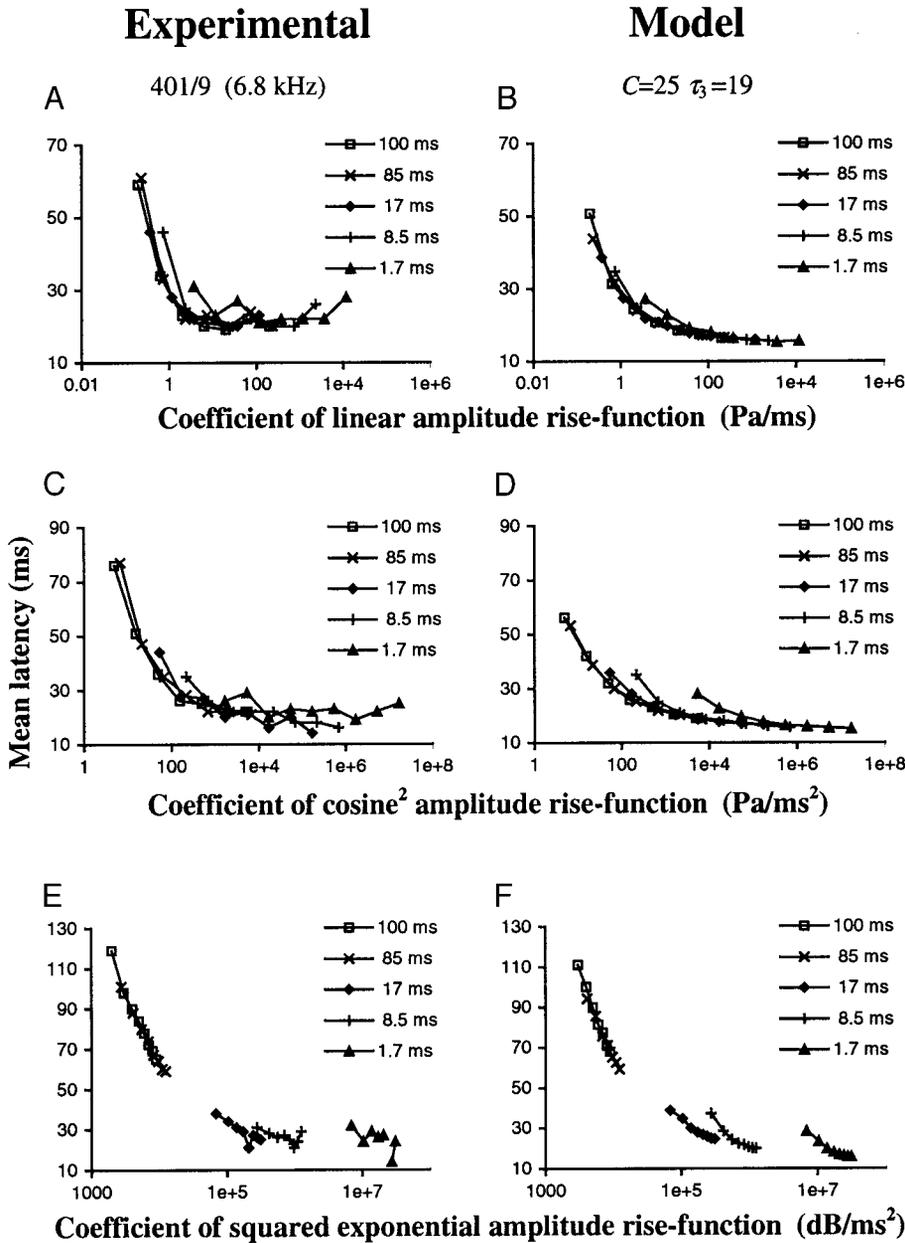


FIG. 7. First-spike latency of a single unit of a cat [primary auditory cortex (AI)] as response to 3 rise functions, linear (A, experimental; B, simulated), cosine squared (C, experimental; D, simulated), and squared exponential (E, experimental; F, simulated). Latency is plotted as a function of the predicted invariant measure of each rise function.

COMPARING MODEL PREDICTIONS WITH LATENCY RESULTS OF PHYSIOLOGICAL EXPERIMENTS. Figure 7 shows the latency data of one AI unit in response to three types of onset windows, linear (Fig. 7A), cosine squared (Fig. 7C), and squared exponential (Fig. 7E). The latency for each window is plotted as a function of the predicted invariant measure, and the alignment of the latency data along a single curve for each rise function validates our predictions. Numerical simulations reproduce these phenomena (Fig. 7, B, D, and F). Figure 8A shows the latency of a MGB neuron in response to four types of amplitude rise function, $\cos^2(t)$, $\cos^4(t)$, t^2 , and t^4 . The latency of each rise function is plotted as a function of the predicted invariant measure, which is P/D^n for the t^n rise functions and $\pi^n P/2^n D^n$ for the $\cos^n(t)$ rise functions {Taylor's series approximation of $\cos^n[(\pi/2)t + (\pi/2)]$ is $(\pi^n/2^n)t^n + o(t^{n+2})$ for even n }. This way the latency data collected with the t^2 and the $\cos^2(t)$ rise function aligns along a single curve, and the latency data collected with the t^4 and the $\cos^4(t)$ rise function

aligns along another curve. The model predictions also hold for the responses of a neuron in primary auditory cortex (Fig. 8C) and are being reproduced by the numerical simulations of the model (Fig. 8, B and D).

SPIKE COUNT. Neurons in AI of anesthetized cat show a low spontaneous rate of fire, and their typical response to sound bursts is a single spike or a short burst of a few spikes immediately following the onset of the stimulus (e.g., Heil 1997b). Examining the spike count as a function of plateau peak pressure alone reveals a non-monotonic pattern that is shared by many AI neurons to various degrees (e.g., Heil 1997b; Heil and Irvine 1998a; Phillips 1988; Schreiner and Mendelson 1990). Furthermore, the non-monotonicity is enhanced at the shorter rise times. Figure 9 demonstrates the typical response patterns of two types of neurons, as replotted from Heil's (1997b) data. Figure 9A shows a highly non-monotonic neuron, whereas Fig. 9C shows a more

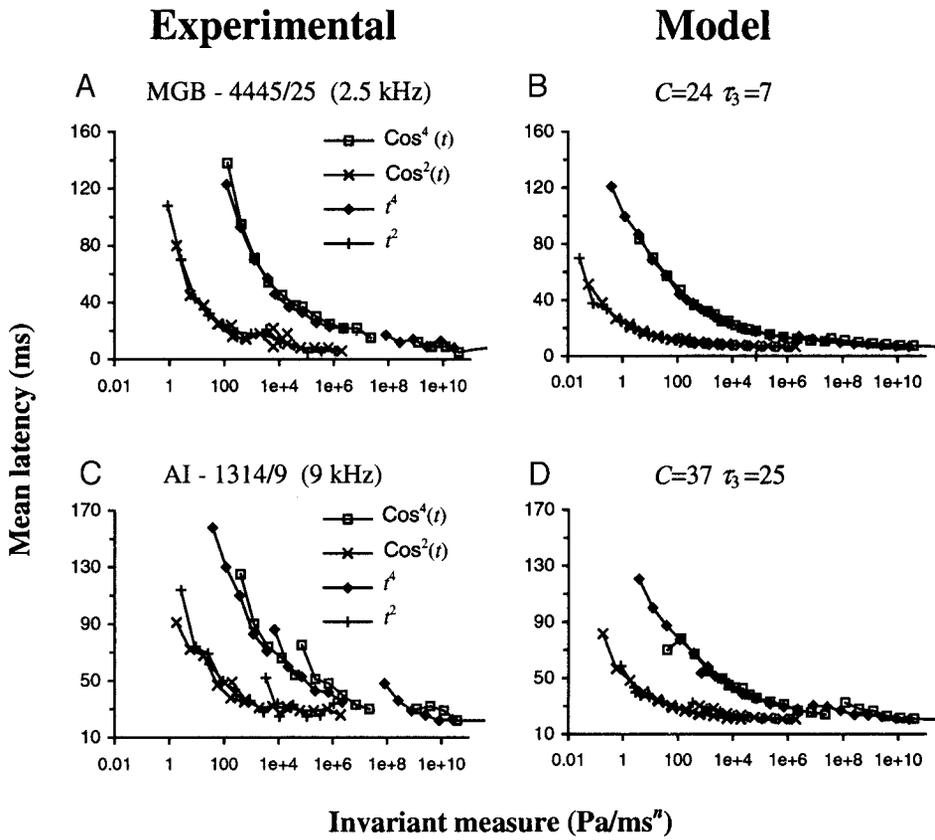


FIG. 8. First-spike latency measured using $\text{cos}^2(t)$, $\text{cos}^4(t)$, t^2 , and t^4 rise functions from a single unit of a cat medial geniculate body (MGB; A and B, simulated) and AI (C and D, simulated). Latency data are plotted as a function of the predicted invariant measure $[P/D^n]$ for the t^n rise functions and $\pi^n P/2^n D^n$ for the $\text{cos}^n(t)$ rise functions].

monotonic neuron. The spike-count data are plotted as a function of plateau peak pressure and are organized along iso-rise-time curves. The model reproduces these phenomena over a wide range of degrees of monotonicity. Figure 9, B and D, demonstrates a good correspondence between

experimental and simulated results. The correspondence is apparent for curve shapes as well as for order of displacement of the iso-rise-time curves, although the displacement of the model curves along the abscissa are much larger than those of the neural curves.

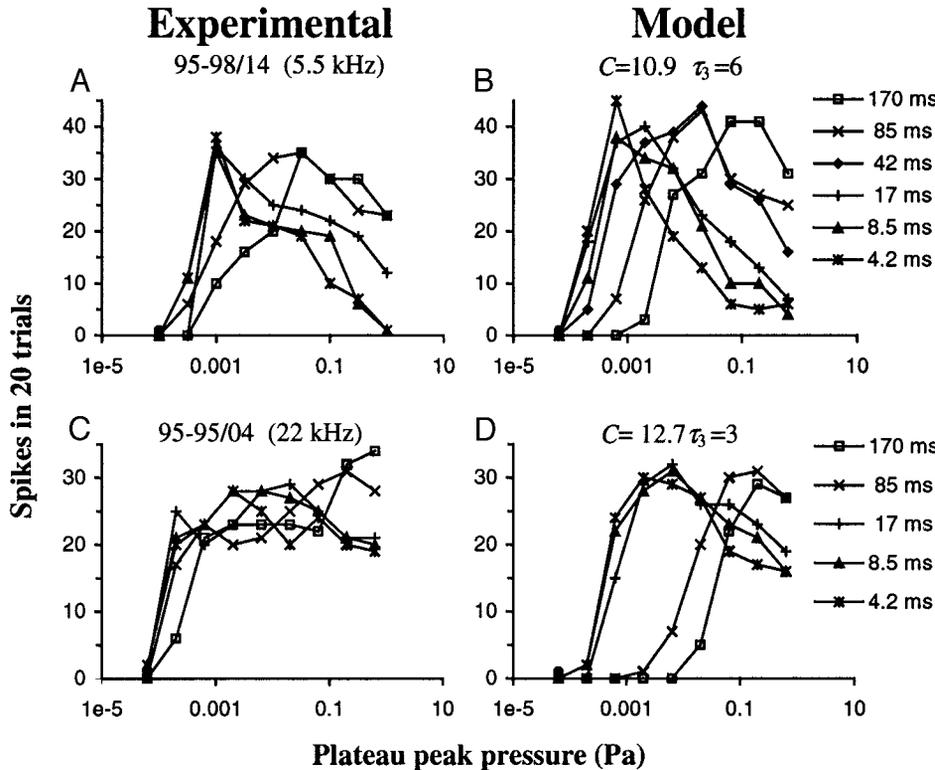


FIG. 9. Experimental vs. simulated spike-count data. Iso-rise-time curves of spike counts are plotted as a function of amplitude level of a cosine-squared onset, for a non-monotonic neuron (A, experimental; B, simulated), and a monotonic neuron (C, experimental; D, simulated). Experimental data are replotted from Heil (1997b). Simulated data for both neurons were obtained using the same sets of parameters that were used to match their latency data (see Fig. 4).

The monotonicity of the model can be controlled by changing the value of the two adjustable parameters, as already illustrated in Fig. 3. Increasing the dynamic range of the delay layer's saturation and decreasing the membrane time constant increase the monotonicity of the neuron. The relation between the degree of the saturation and the monotonicity of the neuron spike count is being formally proved in APPENDIX B. It is noteworthy that raising the sigmoidal scaling factor of the saturation transformation both raises the threshold and increases the monotonicity of the neuron. This relation between the threshold and monotonicity of the neuron is consistent with previously reported findings (Heil et al. 1994; Sutter and Schreiner 1995).

Heil (1997b) found an interesting relationship between the spike count and the latency of the response. This relation links the dynamics of the onset and the number of spikes that follow it. Heil demonstrated that plotting the spike count as a function of the stimuli's peak pressure at the moment of first-spike generation brings the iso-rise-time curves to close congruence. The moment of first-spike generation is defined as the mean latency (for the given rise time and plateau peak pressure) minus the minimal latency of the neuron (as defined by the term L_{\min} in Eqs. 5 and 6). The congruence of the iso-rise-time curves holds for both linear and cosine-squared onsets and for both monotonic and non-monotonic neurons. Figure 10 demonstrates this phenomenon using the data of Heil (1997b), Phillips (1998), and the original data reported here, and shows that the model reproduces this phenomenon for variety sets of model parameters. In APPENDIX B we analyze this special relationship between the latency and the spike count of the model and of auditory cortical neurons.

Evaluation of the neural model: evoked auditory brain stem responses

The ability of the model to match reported evoked auditory brain stem responses in humans (Barth and Burkard 1993) and inferior colliculus potential (ICP) in the awake chinchilla (Phillips and Burkard 1999) in response to sound bursts, was tested by feeding the model with the amplitude envelope of the stimuli and comparing the model output to the reported responses. The membrane potential of our modeled edge detector neuron (Fig. 1E) was used as an estimate of the combined activity of a large population of brain stem neurons (Gerstner 1999b). The model activity was then differentiated to mimic the analogue highpass filter (with a slope of 6 dB/oct) used in these experiments.

Figure 11A shows a typical measure of the inferior colliculus potential in response to a tone burst as replotted from Barth and Burkard (1993). Figure 11B shows the differentiated membrane potential of the edge detector neuron of the model. This figure also illustrates the definitions of the latency and amplitude of the response. The two adjustable parameters that shape the model's response to amplitude transients (C and τ_3) were adjusted to fit the latency and amplitude of the experimental responses.

In contrast to the stimuli that were used in single-cell recordings, whose total durations were 50–100 ms (Phillips 1998) or 400 ms (Heil 1997a,b), the stimuli that were used by Barth and Burkard (1993) and by Phillips and Burkard (1999)

were much shorter and included plateau-level durations of 2–5 ms. For the model to accurately reproduce the experimental responses to these very short bursts, we had to reduce the time constant of the delay layer units from a range of 3–5 ms to a range of 0.5–1 ms, since higher time constants oversmoothed the envelope. The problem of using very short time constants when modeling mammalian inferior colliculus neurons has also been encountered in other modeling studies (Hewitt and Meddis 1994).

LATENCY. Phillips and Burkard (1999) measured the latency of the ICP in the awake chinchilla in response to cosine-squared onsets of various rise times and amplitude levels. Although Phillips and Burkard reported that there were strong similarities between the latency behavior of the ICP and that of cortical single cells, they did not use Heil's functional expression (see Eq. 6) to match the latency data according to the maximum acceleration of the onset envelope. Figure 12A shows that replotting the ICP latency data as a function of maximum acceleration of the envelope brings the iso-rise-time curves to converge along a single curve that can be fitted using Eq. 6 and by using the same value of the constant parameter (A_c) that was used by Heil (1997a). Figure 12B shows that the model reproduces the ICP latency data.

Barth and Burkard (1993) measured the latency of wave V of brain stem auditory evoked responses (BAER) in response to linear shaped onsets. Although Barth and Burkard reported that both the onset rise time and amplitude affect the response latency, they did not analyze the latency as a function of the envelope rate of change. Figure 12C shows that replotting the BAER latency as a function of the envelope rate brings the iso-rise-time curves to close congruence. Using Heil's functional form (Eq. 5) and the same value of Heil's constant (A_l) to match this curve yields a moderate fit. The model latency data is shown in Fig. 12D.

RESPONSE AMPLITUDE. The effect of onset rise time and amplitude level on the ICP and on wave V of BAER response amplitude are similar to their effect on the spike count of monotonic cortical single cells. The response amplitude increased with ascending amplitude levels and with descending onset rise times. Figure 13 replots Phillips and Burkard's (1999) ICP amplitude response (Fig. 13A) and Barth and Burkard's (1993) BAER wave V response amplitude (Fig. 13C); both are plotted as a function of the plateau peak level. The simulated response amplitudes are presented in Fig. 13, B and D, and are scaled in order to match the experimental measurements.

Results of the psychoacoustic experiments

The results of the two experiments were analyzed using a stepwise logistic regression. The dependent variable was set to be the probability of eliciting a discontinuous response, and the independent variables included the stimuli parameters used in each experiment (as detailed in Tables 1 and 2, respectively). In addition, motivated by our model, we added to the two sets of independent variables: the logarithm of the normalized rate of change of the ramp peak pressure re the base peak pressure

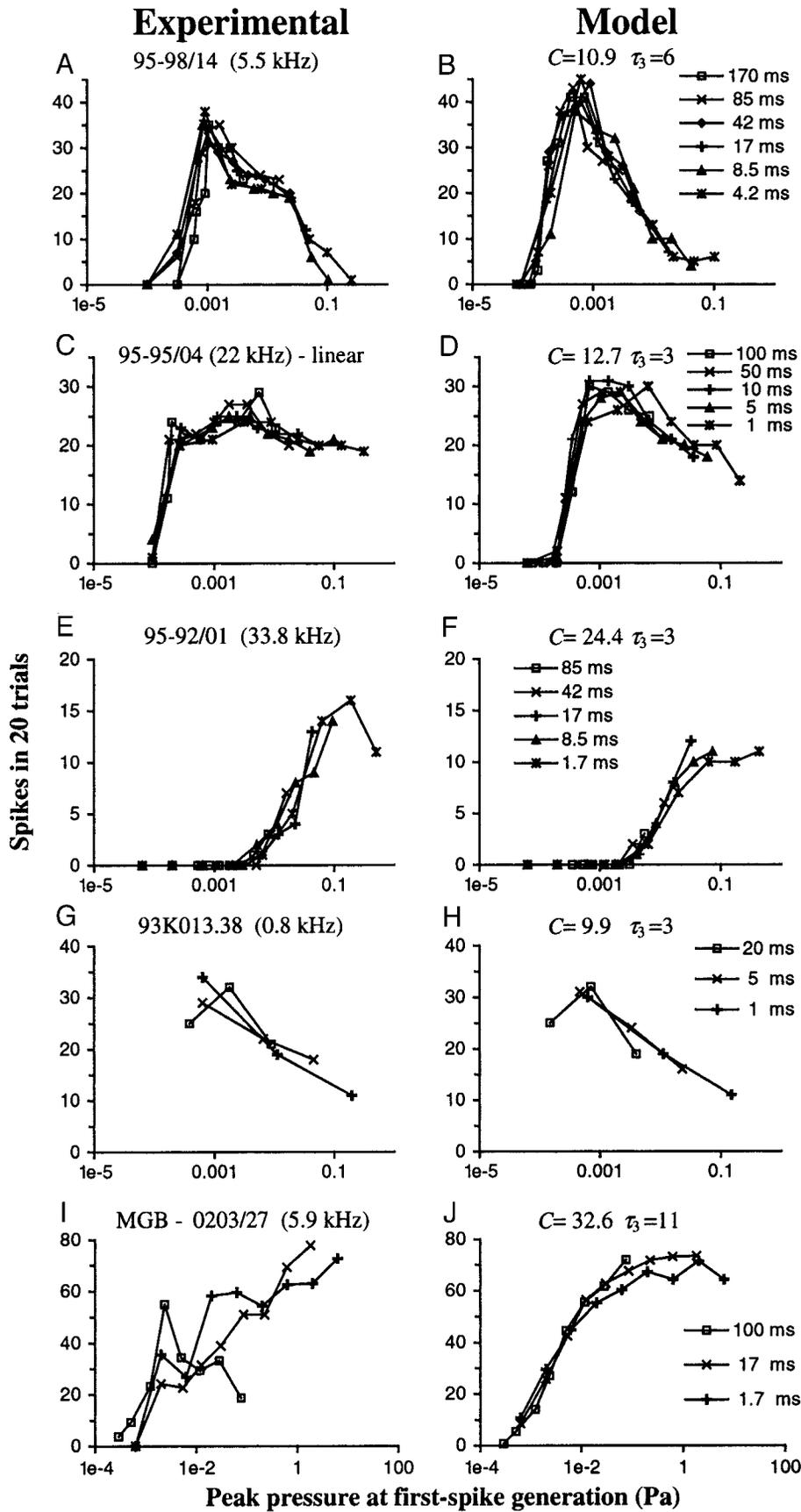


FIG. 10. Experimental and simulated spike-count iso-rise-time curves are closely aligned when plotted as a function of stimulus peak pressure at 1st-spike generation. Experimental data of Heil (1997a,b) (A, C, and E) and of Phillips (1998) (G) are recorded from single units of the cat AI. Original data from a single unit of the cat MGB are shown in I. Note that the model (B, D, F, H, and J) reproduces this phenomenon over a broad range of parameters.

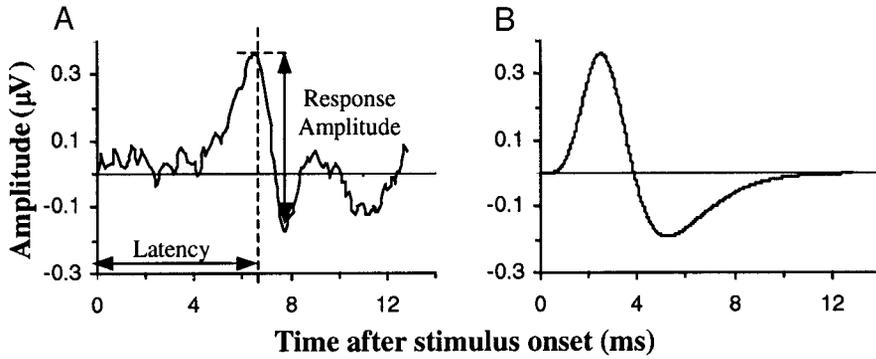


FIG. 11. A: a typical wave-V brain stem auditory evoked response (BAER) as response to a 60-dB nHL 1.25-ms rise-time noise burst replotted from Barth and Burkard (1993). B: a differentiated membrane potential of the model edge detector neuron as a response to the same stimulus without the addition of noise. The response latency is measured with respect to the peak of the BAER, and the response amplitude is measured from the peak to the following trough, as illustrated in A.

(this is the invariant measure for the stimuli used here, see next section).

EXPERIMENT 1. The regression results show that the variable that accounts for most of the variance is the normalized rate of the ramp peak pressure [$F_{(1,2238)} = 1,948.6, P < 10^{-15}$]. Other significant variables are the duration of the change, [$F_{(1,2237)} = 60.8, P < 10^{-12}$]; and the tone frequency [$F_{(1,2236)} = 32.5, P < 10^{-7}$].

EXPERIMENT 2. The results of the second experiment also found the normalized rate of peak pressure to be the variable that accounts for most of the variance [$F_{(1,898)} = 509.6, P < 10^{-15}$]. Other significant variables were the first plateau amplitude level, [$F_{(1,897)} = 29.2, P < 10^{-7}$]; the step amplitude [$F_{(1,896)} = 11.5, P < 0.0007$] and the step duration [$F_{(1,895)} = 11.07, P < 0.0009$].

Mean results across all participants for the two experiments are plotted in Fig. 14A. As expected from the regression analysis, it is evident that plotting the probability data as a function of the rate of peak pressure causes the data to align along a typical psychometric function.

Evaluation of the neural model: psychoacoustic data

In the following section we will compare the model responses with the results of three psychoacoustical experiments. These experiments include the experiment reported above that tested the perception of amplitude discontinuity; an experiment that tested the effect of amplitude transients on auditory segregation (Bregman et al. 1994b); and a forward masking experiment (Turner et al. 1994) that tested the effect of the probe rise time on the degree of masking. Although these experiments investigate different auditory phenomena, we demonstrate that by identifying the psychoacoustical measures with the responses of the neural model to the amplitude transients presented in the experiments, the model is able to reasonably reproduce the psychoacoustical results.

In two of these experiments (Bregman et al. 1994b and the experiment reported here) the stimuli contained an amplitude ramp rising above a pedestal. The invariant measure for these stimuli is not the rate of rise of the amplitude ramp per se, but rather the normalized rate of rise re the pedestal, P^*/D^n , where P^* is the plateau peak pressure of the ramp normalized by the

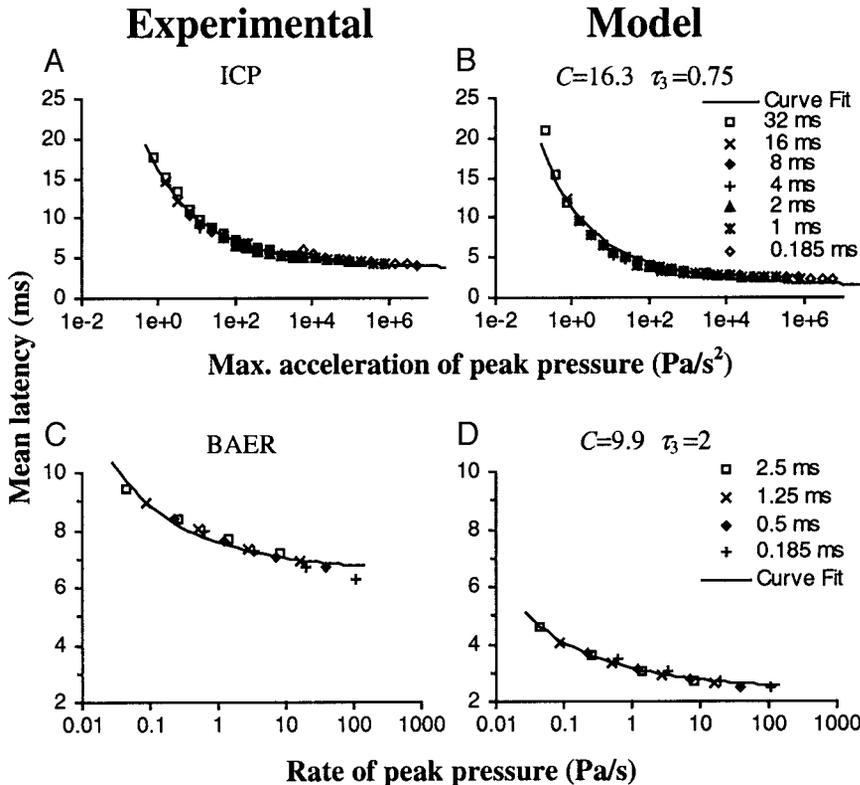


FIG. 12. A: replotting inferior colliculus potential (ICP) response latencies (Phillips and Burkard 1999) as a function of maximum acceleration of cosine-squared onsets yields a good alignment of the latency data along a curve that can be fitted by Heil's (1997a) functional form (Eq. 6). Our fit for the experimental ICP data yields $S = 5.65$ and $L_{\min} = 3.46$. The model reproduces these results (B), with a fit of $S = 5.97$ and $L_{\min} = 1.18$. C: replotting wave-V latencies (Barth and Burkard 1993) as a function of the rate of linear onsets reveals good alignment along a curve that only moderately fits Heil's functional form (Eq. 5) with $S = 5.65$ and $L_{\min} = 6.43$. The model matches the experimental results (D) but is better fitted by the functional form ($S = 6.17$ and $L_{\min} = 2.27$). Note that both Barth and Burkard (1993) and Phillips and Burkard (1999) used 0-ms rise time onsets. To allow a valid calculation of the envelope maximum acceleration and rate of change for these stimuli, we replaced the zero rise time by a 0.185-ms value. This value was found to best match the fitted curves for both the ICP and the BAER latency data.

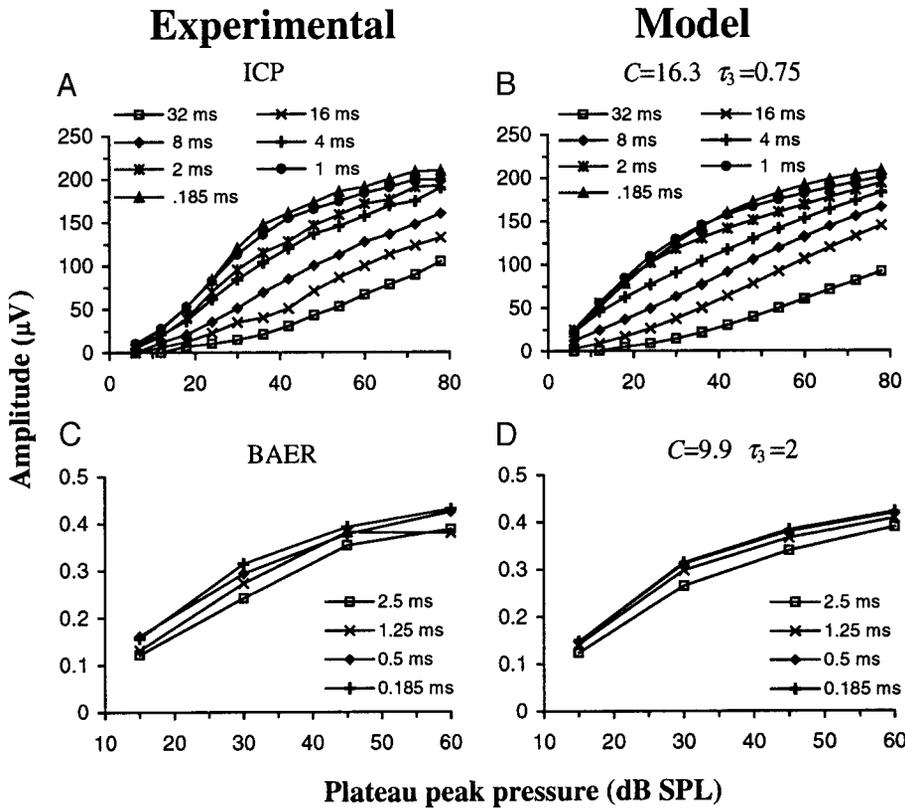


FIG. 13. Response amplitude of ICP (A) replotted from Phillips and Burkard (1999), and response amplitude of wave-V BAER (B) replotted from Barth and Burkard (1993) as a function of plateau peak pressure. Note the resemblance between the 2 experimental findings in response to stimuli of comparable parameters and between the experimental and simulated data (B and D).

ratio between the pedestal peak pressure and P_0 (see Fig. 1B). Intuitively, this follows from the fact that the essential operation of the model is differentiating the log-compressed amplitude envelope. Therefore the output of the receptive field (Fig. 1D) is not changed by multiplying the input stimuli by a constant factor. In consequence, the response to a ramp rising

above a pedestal is identical to the response to the onset of a sound with the same size in dB re P_0 and with the same shape. Note that we arbitrarily set the value of P_0 to 0 dB SPL for simplicity sake. Using different P_0 values can be compensated by adjusting the threshold value of the edge detector neuron. P_0 value is significant only when fitting the model responses with

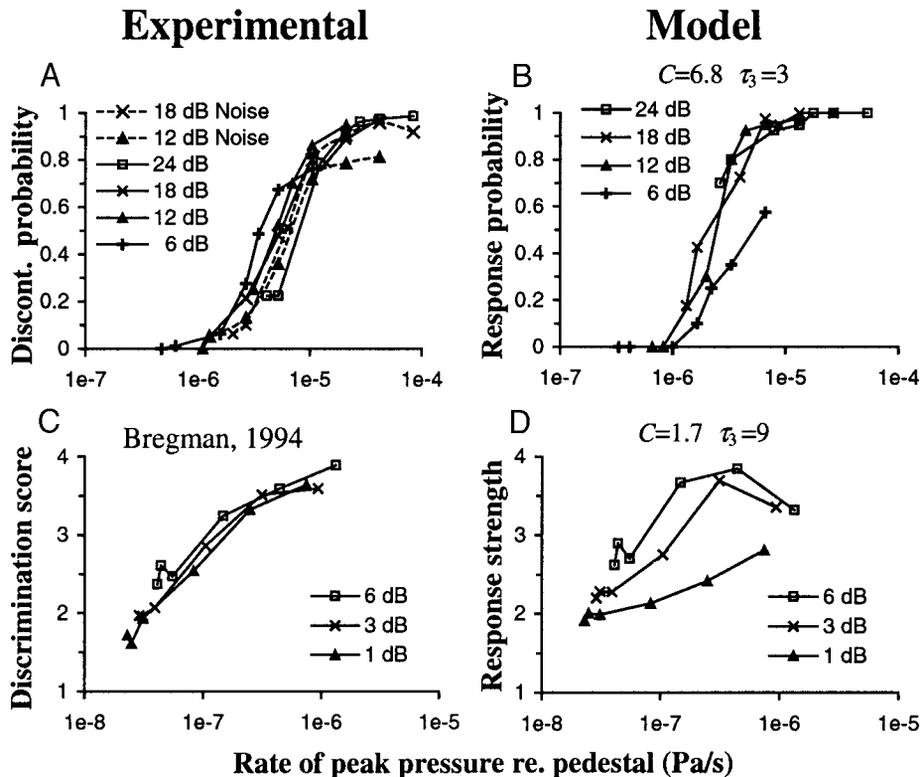


FIG. 14. A: mean results across all participants for *experiment 1* (solid lines) and *experiment 2* (dashed lines). The probability for the amplitude ramp to be perceived as a discontinuous change is plotted as a function of the normalized rate of the ramp peak pressure re the pedestal (see text). This produces a good congruence of the data along a typical psychometric curve. A plot of the simulation results (B) shows a good fit with the psychoacoustic data. C: a replot of the discrimination score from Bregman et al. (1994b) as a function of the normalized rate of change of the incremented partials. The model matches the data only moderately (D).

the responses of a specific neuron to both the onset of a sound and to a ramp rising above a pedestal. In these cases P_0 may be adjusted to best fit the neural responses to both types of stimuli.

PERCEPTION OF AMPLITUDE DISCONTINUITY. To compare our psychophysical results and the model predictions, a function of the neural response compatible with the dichotic nature of the psychophysical responses is required. As mentioned earlier, the modeled neurons have low spontaneous activity, and their responses to sound bursts consist of a short burst of 1–3 spikes. Therefore it seemed plausible to define a response to a stimulus as one or more spikes, and to identify the probability of response as the probability that a participant would report a discontinuous amplitude change in the psychophysical experiment. This measure did in fact yield a good match between the simulated (Fig. 14B) and the experimental results.

EFFECT OF AMPLITUDE TRANSIENTS ON AUDITORY SEGREGATION. One of the few studies that tested the effect of both the duration and magnitude of amplitude changes on auditory segregation tasks has been reported in Bregman et al. (1994b). They presented a 3.5-s long complex tone consisting of five harmonics of 500 Hz. The amplitudes of an adjacent pair of the three middle frequencies (1,000, 1,500, and 2,000 Hz) were incremented in succession in random order. A sufficiently large amplitude increment caused the partials to be segregated from the complex tone, and to be perceived as separate tones. To measure the degree of segregation, the participants had to judge whether the perceived pitch pattern, caused by the segregated partials, went up or down. Three levels of increments were used (1, 3, and 6 dB) and six increment durations (30, 90, 270, 730, 910, and 970), resulting in a total of 18 experimental conditions. The overall amplitude level of the complex tone in its steady state was 65 dB SPL. Bregman et al. reported that both the amplitude increment level and the increment duration had a significant effect on the participants' performance. Longer increment duration resulted in poorer discrimination performance, while larger increment levels led to better discrimination. These results suggest that the gradient of the increment had a dominant effect on discrimination performance. However, Bregman et al. did not include the gradient of the increment in their statistical analysis, and therefore it is impossible to determine the exact influence of the amplitude gradient of a tone on the ability to segregate it from a mixture of tones. When the results of Bregman et al. are replotted as a function of the normalized rate of peak pressure of the amplitude increment, the data fall along a single curve (Fig. 14C).

Since Bregman et al. used a continuous measure ranging from 0 to 5, we used the spike count of the model as the simulated measure while using a linear transformation of the spike count data that resulted in the best fit to the psychoacoustical results. Figure 14D shows that the model's ability to approximate the experimental results of Bregman et al. is only moderate. Formally, the model responses do not align on a single curve because of the use of extremely shallow ramps in this experiment (see Fig. B1 and the accompanying discussion in APPENDIX B). Interestingly, the experimental data of Bregman et al. (1994b) are in fact invariant with respect to the normalized rate of rise of the ramp, implying that the rate of rise is the behaviorally relevant variable even under these extreme conditions.

EFFECT OF AMPLITUDE TRANSIENTS ON RELEASE FROM FORWARD MASKING. In forward masking, the masker (which can be a tone or a noise burst) masks a target tone that appears just after the masker ends. The degree of masking depends on many factors such as masker level, bandwidth, duration, and the inter-stimulus interval. Turner et al. (1994) studied the effect of the target tone rise time and duration on forward masking levels. They used two types of target tones, one with a total duration of 25 ms including 2-ms cosine-squared rise/fall ramps, and the second with a total duration of 22 ms including 10-ms cosine-squared rise/fall ramps. Growth of masking (GOM) functions were measured using noise maskers at levels of 10–90 dB SPL. Their results show that targets with 10-ms rise time were masked more than targets of 2-ms rise time. In addition, Turner et al. showed that in contrast with the psychoacoustical results, there was no significant effect of the target rise time on the amount of masking that was measured in single auditory-nerve fibers of the chinchilla. This suggests that, although some forward masking effects are apparent at the level of the auditory periphery, the effect of target rise time may involve higher auditory centers.

To put these results in the context of our model, we interpreted the forward masking paradigm as a method of assessing the strength of response produced by the target tone; the higher the response produced by the target, the louder the masker that is needed to mask it. Therefore we interpreted the minimal masker level needed to mask a target tone as a measure of the response produced by the target. This measure is being compared with the strength of response produced by the neural model as response to the target tone alone. Figure 15A replots the masker level as a function of the target level for the two rise-time targets as calculated from the data of Turner et al. Figure 15B demonstrates that these results are reproduced by the spiking responses of the edge detector neuron in the model. In addition, the data of Turner et al. remarkably resemble the ICP amplitude data of Phillips and Burkard (1999). Figure 15C replots Phillips and Burkard's (1999) ICP responses at comparable parameter values, and the corresponding model responses (as already shown in Fig. 13) are plotted in Fig. 15D. Thus the psychophysical data of Turner et al. can also be interpreted by this version of the model.

DISCUSSION

In the present study we describe a neural model for auditory temporal edge detection. The core of the model is in the formation of an auditory delay dimension. Sensitivity to amplitude edges is achieved by differentiating the stimulus along this dimension. We demonstrate the ability of the model to reproduce both the latency and magnitude of responses to sound bursts, as recorded from single units of the cat primary auditory cortex and posterior field (Heil 1997a,b; Heil and Irvine 1996; Phillips 1988, 1998), inferior colliculus potential of awake chinchilla (Phillips and Burkard 1999), and wave V of human brain stem-evoked response (Barth and Burkard 1993). Moreover, we predict the response of cortical neurons to a general family of sound bursts whose onset envelope is a power function or the exponent of a power function. We successfully verified these predictions for several of these stimuli by recording from single units of the cat primary auditory cortex and MGB.

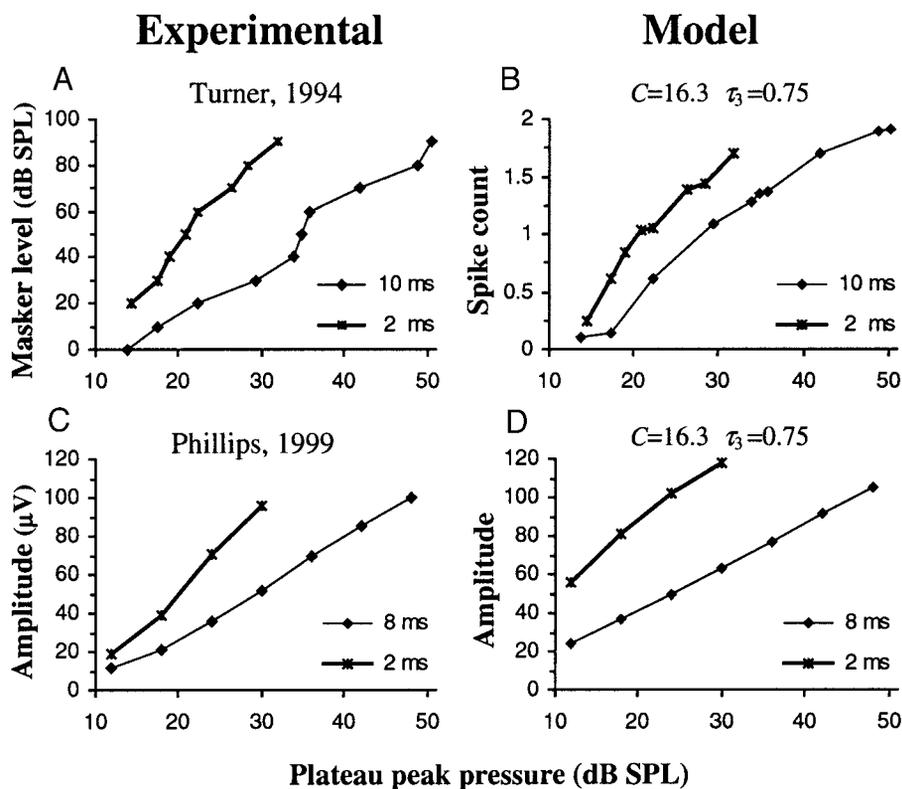


FIG. 15. A: forward masking data from Turner et al. (1994). The minimal noise masker level needed to mask a target tone is plotted as a function of the target level. It is obvious that a more intense masker is needed for targets with higher intensities and shorter rise times. C: the results of Turner et al. show great similarity to the experimental (C) and simulated (D) ICP level of response to comparable noise bursts (Phillips and Burkard 1999), as was plotted in Fig. 13. B demonstrates that the model also reproduces the data of Turner et al. when a spike-count measure is used instead of the differentiated membrane potential measure as in D.

In addition, we tested the ability of the model to match psychoacoustical findings for the sensitivity of human perception to amplitude transients. Our results show that the model is capable of reproducing psychoacoustical results for the effect of amplitude gradient on auditory segregation (Bregman et al. 1994b); the effect of amplitude gradient on the ability to release a tone from a forward masker (Turner et al. 1994); and the effect of amplitude gradient on the perception of the amplitude transient itself as measured in the experiments reported here. The behavior of the model stems from its general operational principles and does not depend on the exact implementation or parameters of any of its components. This important property of the model is established by a mathematical analysis of the model's operation.

Although the model usually follows the experimental data very accurately, there is one prominent systematic deviation of the simulated results from the experimental results. This deviation occurs at relatively long rise times at near threshold levels of plateau peak pressure. In these conditions the model spike count and latency are smaller than the experimental ones (see Figs. 4 and 9 for latency and spike-count data, respectively). This deviation causes Heil's fit for the latency data to produce higher S values for the simulated data than for the experimental data. However, the underestimation of both latency and spike count in the simulated responses preserves the special latency-spike count relationships, in line with the experimental data (Fig. 10). The same effect causes the fit of the model to the data of Bregman et al. (1994b) to be rather poor.

While all the elements of the model are simple and biologically plausible, the use of an auditory delay layer currently lacks definite physiological or anatomical evidence. However, there is some evidence that may validate the use of such an auditory delay layer. Hattori and Suga (1997) measured the

latency of single and multiple neurons from the inferior colliculus (IC) of unanesthetized mustached bats as a response to tone bursts. They found that the latency (ranging from 4 to 12 ms) is topographically organized orthogonally to the tonotopic organization of the IC, forming a frequency versus latency map. Similar organization of onset latencies in the cat IC was reported by Schreiner and Langner (1988b). They reported that the latency of response to CF tones at 60 dB above threshold (ranging from 5 to 18 ms) systematically varied across a given frequency band lamina. Both the range of values and the topographic organization of the latency in the bat and in the cat IC are consistent with the model's delay layer. However, more research is needed to establish a direct link between these findings and the proposed model. Some organization of minimal latency along the isofrequency contours is also present in cat auditory cortex (Mendelson et al. 1997), possibly reflecting a similar map in the cat IC.

The main contribution of the proposed model lies in its ability to reproduce diverse physiological and psychophysical findings on the sensitivity of the auditory system to amplitude transients, especially since currently there is no theoretical framework to which these experimental phenomena can be associated. The motivation for our study stems from the conjecture that auditory transients could supply important cues for the perceptual task of auditory source separation. The problem of sensory source separation is an extremely difficult one, especially when the input contains information that originates from an unknown number of sensory sources of unknown type and location. Since the solution space for almost any given input is infinite, some assumptions regarding the nature of the input need to be made. One basic assumption that is believed to be used by the visual system is that the brightness gradient within an object cannot be too large. This implies that when-

ever a sudden brightness change (visual edge) is observed, it is interpreted as a border between adjacent objects. The existence of neurons in the visual system that are sensitive to brightness edges supports the conjecture that the visual system uses local gradient constraints when interpreting visual images.

This visual example of a priori constraints that reduce the solution space for the source separation problem led us to make two assumptions that underlie the work presented here. First, we assume that the local gradient constraint can be applied to the perception process of acoustic signals. Second, we assume that local gradients of acoustic properties can be computed using neural circuitry that is similar to the one that is used to compute local gradients of visual properties in sub-cortical visual centers. These assumptions lead to two expectations.

First, we would expect to find units of the auditory system that are sensitive to the gradient of the stimulus amplitude. Indeed, as reviewed earlier, examination of the responses of many cortical and sub-cortical neurons to amplitude transients suggests that the neural response is sensitive to the derivative of the stimulus intensity over time and therefore their response may be interpreted as reflecting a temporal edge detection computation.

Second, we would expect to find that amplitude gradients affect auditory perception in general and auditory source segregation phenomena in particular. Although many studies demonstrate the importance of amplitude transients to speech intelligibility (Drullman et al. 1994a,b; Shannon et al. 1995) and to the segregation process of a sinusoidal component from a background of other sinusoidal tones (Bregman et al. 1994a), the importance of the amplitude gradient cannot be directly deduced from these observations. Only few psychophysical studies (Bregman et al. 1994b; Turner et al. 1994) have explicitly manipulated both the duration and the size of the amplitude change simultaneously, making it possible to isolate the effect of the amplitude gradient on auditory perception. As we have demonstrated earlier, the results of these studies are consistent with the assumption that auditory perception is sensitive to the gradient of amplitude transients and that a larger gradient enables easier separation of auditory components.

An alternative explanation for these physiological and psychoacoustical phenomena is that they reflect the sensitivity of the auditory system to the frequency splatter that may be caused by an amplitude transient, rather than by the transient per se. However, this explanation is rendered implausible by many experiments that demonstrate the effect of amplitude transients using broad-band noise bursts (e.g., Barth and Burkard 1993; Phillips and Burkard 1999; Turner et al. 1994; and the psychoacoustical experiments reported here).

These physiological and psychophysical findings support our assumption that the local gradient constraint may be applied to the perception process of acoustic signals. These observations, and the assumption regarding the possible similarity between neural mechanisms that perform visual and auditory edge calculations, led us to suggest the proposed model whose underlying principles are inspired by classical models for visual edge detection neurons.

The ability of the model to account for numerous disparate experimental findings suggests that the sensitivity of the auditory system to amplitude transients is a realization of auditory temporal edge calculation, and that this computation has a

primary role in neural auditory processing in general and in auditory source separation in particular.

APPENDIX A

This section lists the mathematical equations and parameters of the model.

Neural representation

The amplitude envelope, $E(t)$, of the input stimulus is logarithmically compressed and low-pass filtered. When expressed in dB SPL units, the neural representation is

$$N(t) = \frac{1}{\tau_1^2} \int_0^t E(x)(t-x)e^{-(t-x)/\tau_1} dx \quad (A1)$$

where τ_1 is set to 1 ms.

Delay layer

The operation of each unit $U(t, \eta_i)$ of the delay layer on its input, $N(t)$ is given by

$$U(t, \eta_i) = \frac{1}{\eta_i^2} \int_0^t N(x)(t-x)e^{-(t-x)/\eta_i} dx \quad (A2)$$

In our simulations we used 10 units with η_i values equally spaced between 3 and 5 ms. The output of the units is saturated using the following sigmoidal transformation

$$\tilde{U}(t, \eta_i) = F_{\max} \{2/[1 + e^{-U(t, \eta_i)C}] - 1\} \quad (A3)$$

where F_{\max} is set to 225 spikes/s and C is a scaling parameter that is used to adjust the degree of the spike-count monotonicity.

Receptive field and edge detection neuron

The delay layer units are connected to a single neuron. The neuron's input $I(t)$ is given by

$$I(t) = \sum_i W_i \tilde{U}(t, \eta_i) \quad (A4)$$

where $W_i = \{0.0285, 0.1637, 0.5240, 0.8547, 0.4697, -0.4697, -0.8547, -0.5240, -0.1637, -0.0285\}$. The neuron is modeled as a simple leaky integrator with a voltage threshold (T), with an absolute refractoriness period $\delta^{\text{abs}} = 1$ ms, and a refractoriness function

$$\varphi(t) = -Te^{-(t-\delta^{\text{abs}})/\lambda} S(t - \delta^{\text{abs}}) - KS(t)S(\delta^{\text{abs}} - t) \quad (A5)$$

with a constant $K \rightarrow \infty$, $\lambda = 1.5$ ms and where $S(t)$ is the positive step function (Gerstner 1999a). The membrane potential $M(t)$ of the neuron is given by

$$M(t) = \frac{1}{\tau_3^2} \int_0^t [I(x) + \xi(x)](t-x)e^{-(t-x)/\tau_3} dx + \sum_{f_i \in \{f_1, \dots, f_n\}} \varphi(t - f_i) \quad (A6)$$

where $\{f_1, \dots, f_n\}$ are the set of firing times of the neuron, the membrane time constant τ_3 is a parameter and $\xi(x)$ is a random gaussian noise with a zero mean and a standard deviation $\sigma = 0.2T$.

APPENDIX B

Approximate expressions for the model components

In the following we derive approximate expressions for the operation of each of the model components on its input, as annotated in Fig. 1. These expressions will be used throughout the appendix for

analyzing some properties of the model. In our analysis we assume that the amplitude envelope of the input stimulus onset, $E(t)$, can be approximated by a power function

$$E(t) = P \left(\frac{t}{D} \right)^n \quad \text{for } 0 \leq t \leq D \quad (\text{B1})$$

where $n > 0$, P denotes the plateau peak pressure in Pascal units and D denotes the onset rise time in milliseconds. For now the analysis will be restricted to $t \leq D$. The implications of this restriction for the analysis of the first-spike latency phenomena have been discussed in RESULTS, and the implications for the analysis of the spike-count phenomena will be discussed in the following. The neural representation of the auditory input is achieved by low-pass filtering of the stimulus envelope in dB SPL units

$$N(t) = \frac{A}{\tau_1^2} \int_0^t \ln \left[1 + \frac{P}{P_0} \left(\frac{x}{D} \right)^n \right] (t-x) e^{-(t-x)/\tau_1} dx \quad (\text{B2})$$

where $A = 20/\ln(10)$ and $P_0 = 2e - 5$ Pa. The form of the argument to the log transformation eases the analysis for near-zero values of t , and has negligible effect for $t \gg 0$.

The convolution integrals appearing at three levels of the model (neural representation, delay layer, and edge detector neuron) do not have closed analytical form. In the following, these integrals are approximated as follows

$$\frac{1}{\tau^2} \int_0^t F(x)(t-x) e^{-(t-x)/\tau} dx \cong \frac{1}{\tau} F(t) \int_0^t (t-x) e^{-(t-x)/\tau} dx \quad (\text{B3})$$

This approximation is valid when $F(x)$ is monotonic increasing, as it is in all three cases, and when τ is small enough so that $F(x)$ varies slowly on an interval comparable to τ around time t . These claims will be proved at the end of this section.

Using this approximation for the neural representation gives rise to the following expression

$$\begin{aligned} N(t) &\cong \frac{A}{\tau_1^2} \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] \int_0^t (t-x) e^{-(t-x)/\tau_1} dx \\ &= A \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] \left[1 - e^{-t/\tau_1} \left(1 + \frac{t}{\tau_1} \right) \right] \end{aligned} \quad (\text{B4})$$

The output of each unit of the delay layer is given by

$$U(t, \tau_2) = \frac{1}{\tau_2^2} \int_0^t N(x)(t-x) e^{-(t-x)/\tau_2} dx$$

substituting the approximation of $N(x)$ according to Eq. B4 yields

$$U(t, \tau_2) \cong \frac{A}{\tau_2^2} \int_0^t \ln \left[1 + \frac{P}{P_0} \left(\frac{x}{D} \right)^n \right] \left[1 - e^{-x/\tau_1} \left(1 + \frac{x}{\tau_1} \right) \right] (t-x) e^{-(t-x)/\tau_2} dx$$

which is approximated by

$$\begin{aligned} U(t, \tau_2) &\cong A \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] \left\{ 1 - e^{-t/\tau_2} \left[\frac{\tau_2^2(\tau_2 - 3\tau_1)}{(\tau_2 - \tau_1)^3} + t \frac{\tau_2}{(\tau_2 - \tau_1)^2} \right] \right. \\ &\quad \left. + e^{-t/\tau_1} \left[\frac{\tau_1^2(\tau_1 - 3\tau_2)}{(\tau_2 - \tau_1)^3} + t \frac{\tau_1}{(\tau_2 - \tau_1)^2} \right] \right\} \end{aligned} \quad (\text{B5})$$

Biological constraints require that $\tau_2 \gg \tau_1 = 1$ ms. In this case Eq. B5 can be further simplified to

$$U(t, \tau_2) \cong A \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] \left[1 - e^{-t/\tau_2} \left(1 + \frac{t}{\tau_2} \right) \right] \quad (\text{B6})$$

The operation of the receptive field is approximated by the negative of the first-order derivative of $\tilde{U}(t, \tau_2)$ with respect to τ_2 at the neighborhood of some fixed value of τ_2 . For the analytical treatment, the sigmoidal saturation following the delay layer will be neglected (the effects of saturation will be analyzed below). The degree of non-monotonicity of the model depends on this transformation. Thus, only monotonic neurons will be further analyzed using these approximations. Using these assumptions, it follows that

$$R(t) = \frac{d\tilde{U}(t, \tau_2)}{d\tau_2} \cong \frac{A}{\tau_2^2} \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] t^2 e^{-t/\tau_2} \quad (\text{B7})$$

The membrane potential of the edge detector neuron is therefore given by

$$M(t) = \frac{A}{\tau_2^2 \tau_3^2} \int_0^t \ln \left[1 + \frac{P}{P_0} \left(\frac{x}{D} \right)^n \right] x^2 e^{-x/\tau_2} (t-x) e^{-(t-x)/\tau_3} dx$$

Given the realistic biological values for τ_2 and τ_3 time constants we can assume that $\tau_2 = \tau_3$ and by using the same kind of approximation used in Eq. B4 we get the following simplification

$$M(t) \cong \frac{A}{12\tau_3^3} \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] t^4 e^{-t/\tau_3} \quad (\text{B8})$$

$M(t)$ will be used to investigate the dependence of the first-spike latency and spike count of the model on the parameters (P , D , and n) of the input stimulus.

To justify the approximation described in Eq. B3, we define $G(t)$ and $\tilde{G}(t)$ as follows

$$G(t) = \frac{1}{\tau^2} \int_0^t F(x)(t-x) e^{-(t-x)/\tau} dx \cong \tilde{G}(t) = \frac{1}{\tau^2} F(t) \int_0^t (t-x) e^{-(t-x)/\tau} dx \quad (\text{B9})$$

where

$$F(x) = A \ln \left[1 + \frac{P}{P_0} \left(\frac{x}{D} \right)^n \right]$$

and investigate the relative error $[\tilde{G}(t) - G(t)]/\tilde{G}(t)$. Since

$$\begin{aligned} \tilde{G}(t) - G(t) &= \int_0^{t_0} [F(t_0) - F(x)](t-x) e^{-(t-x)/\tau} dx \\ &\quad + \int_{t_0}^t [F(t) - F(x)](t-x) e^{-(t-x)/\tau} dx \quad 0 \leq t_0 \leq t \end{aligned}$$

and since $(t-x)e^{-(t-x)/\tau} \geq 0$ for $0 \leq x \leq t$, and $F(x)$ is positive and monotonically ascending, it holds that

$$\begin{aligned} \tilde{G}(t) - G(t) &\leq [F(t_0) - F(0)] \int_0^{t_0} (t-x) e^{-(t-x)/\tau} dx \\ &\quad + [F(t) - F(t_0)] \int_{t_0}^t (t-x) e^{-(t-x)/\tau} dx \end{aligned}$$

and therefore

$$\begin{aligned} \frac{\tilde{G}(t) - G(t)}{\tilde{G}(t)} &\leq L(t_0, t) = \frac{1 - E(t - t_0)}{1 - E(t)} - \frac{[F(t_0)]}{[F(t)]} \frac{1 - 2E(t - t_0) + E(t)}{1 - E(t)} \\ &\quad \text{where } E(t) = e^{-t/\tau} \left(1 + \frac{t}{\tau} \right) \end{aligned}$$

It is easy to see that for a fixed value of t , there is \hat{t}_0 such that $L(\hat{t}_0, t) \leq L(t_0, t)$ for every $0 \leq t_0 < t$. Although we are unable to show a closed analytic form for \hat{t}_0 , there are some observations that can be made regarding $L(t_0, t)$. If $F(x)$ is a constant function such that $F(t_0)/F(t) =$

1 for every $t_0 < t$ then $\hat{t}_0 = 0$ and $L(\hat{t}_0, t) = 0$. Given the function $F(x)$ that is being used in Eq. B9, it holds that

$$\lim_{P \rightarrow \infty} \frac{F(t_0)}{F(t)} = 1 \quad \text{for every } 0 < t_0 < t \leq D \Rightarrow \lim_{P \rightarrow \infty} L(t_0, t) = 0$$

and

$$\lim_{n \rightarrow 0} \frac{F(t_0)}{F(t)} = 1 \quad \text{for every } 0 < t_0 < t \leq D \Rightarrow \lim_{n \rightarrow 0} L(t_0, t) = 0$$

On the other hand, $\lim_{D \rightarrow 0} F(t_0)/F(t)$ and $\lim_{D \rightarrow \infty} F(t_0)/F(t)$ for every $0 < t_0 < t \leq D$ depend on the exact form of F . In the special case that $D \rightarrow \infty$ or $\tau \rightarrow 0$, then $t_0/\tau, t/\tau \rightarrow \infty$, and it holds that

$$\lim_{t_0/\tau, t/\tau \rightarrow \infty} L(t_0, t) = 1 - \frac{F(t_0)}{F(t)} \Rightarrow \hat{t}_0 = t \Rightarrow \lim_{t_0/\tau, t/\tau \rightarrow \infty} L(\hat{t}_0, t) = 0$$

and when $D \rightarrow 0$ or $\tau \rightarrow \infty$, then $t_0/\tau, t/\tau \rightarrow 0$, and it holds that

$$\lim_{t_0/\tau, t/\tau \rightarrow 0} L(t_0, t) = \frac{(a-1)^2}{a^2} - \frac{(a-2)^2 - 2F(t_0)}{a^2 F(t)} \quad \text{where } a = \frac{t}{t_0}$$

which is always positive.

These observations show that the quality of the fit becomes better as P and D becomes bigger and n and τ becomes smaller. For example, if $F(t)$ varies slowly enough such that $F(\tau)/F(t) > 0.95$ for $\tau < t \leq D$ then

$$\frac{\tilde{G}(t) - G(t)}{\tilde{G}(t)} \leq 0.157 \quad \text{for } t > 4\tau, \quad n = 1$$

Proof for the compressive nature of the model's latency as a function of the rise time

As illustrated in Fig. 3G, it is assumed that the edge detector neuron starts firing when its membrane potential hits a fixed threshold level, T . Thus the time of the first spike, t^* , satisfies the condition: $M(t^*) = T$, where $M(x)$ is the edge detector membrane potential approximated by Eq. B8. To prove the compressive nature of $t^*(D)$, we will show that $\lim_{D \rightarrow 0} (d/dD)t^*(D) = \infty$ and $\lim_{D \rightarrow \infty} t^*(D) < \infty$. Since $\lim_{D \rightarrow 0} t^*(D) = 0$ and $t^*(D) > 0$ for every D , the compressive nature of $t^*(D)$ is proven.

To prove that $\lim_{D \rightarrow 0} (d/dD)t^*(D) = \infty$, we will differentiate both sides of the following approximation with respect to D

$$\frac{d}{dD} [M(t^*)] \cong \frac{d}{dD} \left\{ \frac{A}{12\tau_3} \ln \left[1 + \frac{P}{P_0} \left(\frac{t^*}{D} \right)^n \right] t^{*4} e^{-t^*/\tau_3} \right\} = \frac{d}{dD} T = 0$$

To extract the following expression for $(d/dD)t^*$

$$\frac{d}{dD} t^* = \frac{\tau_3 n [e^{g(t^*)} - 1]^{1/n}}{\frac{P}{P_0} \left\{ [\tau_3 n + 4\tau_3 g(t^*) - t^* g'(t^*)] + \frac{g(t^*)}{e^{g(t^*)} - 1} (4\tau_3 - t^*) \right\}}$$

$$\text{where } g(t^*) = \frac{12T\tau_3^5}{A t^{*4} e^{-t^*/\tau_3}}$$

Since $\lim_{D \rightarrow 0} t^*(D) = 0$ and $\lim_{t^* \rightarrow 0} g(t^*) = \infty$ it holds that

$$\lim_{D, t^* \rightarrow 0} \frac{d}{dD} t^* = \lim_{D, t^* \rightarrow 0} \frac{\tau_3 n \frac{1}{g(t^*)} [e^{g(t^*)} - 1]^{1/n}}{\frac{P}{P_0} \left\{ \left[\frac{\tau_3 n}{g(t^*)} + 4\tau_3 - t^* \right] + \frac{1}{e^{g(t^*)} - 1} (4\tau_3 - t^*) \right\}} = \infty$$

In order to prove that $\lim_{D \rightarrow \infty} t^*(D) < \infty$ the following function is defined

$$Z(D) = \hat{t} \quad \text{such that} \quad M(\hat{t}) = \max_{t > 0} \{M(t)\}$$

$Z(D)$ is the time-to-peak of the membrane potential. It will be shown below that

$$\lim_{D \rightarrow \infty} Z(D) = (4+n)\tau_3 \quad (\text{B10})$$

From the definition of $Z(D)$ it holds that

$$\lim_{D \rightarrow \infty} t^*(D) \leq \lim_{D \rightarrow \infty} Z(D) = (4+n)\tau_3$$

To prove Eq. B10, we solve the following equation

$$\lim_{D \rightarrow \infty} \frac{d}{dt} M(t) = 0 \Rightarrow \lim_{D \rightarrow \infty} \left\{ \tau_3 n \frac{P}{P_0} t^{n+3} + D^n \ln \left[1 + \frac{P}{P_0} \left(\frac{t}{D} \right)^n \right] (4\tau_3 t^3 - t^4) \right\} = 0$$

Dividing both sides of the equation by D^n yields

$$\frac{P}{P_0} t^{n+3} [\tau_3(n+4)] - \frac{P}{P_0} t^{n+4} = 0 \Rightarrow \begin{cases} t = 0 \\ t = \tau_3(n+4) \end{cases}$$

The non-trivial solution of the equation proves Eq. B10. Note that these are asymptotic results; numerical analysis of $M(t)$ demonstrates that for finite values of D , the derivative of $t^*(D)$ with respect to D is not proportional to $1/P$, in agreement with the experimental findings of Heil and Irvine (1996) and of Phillips (1998).

Proof for the effect of the saturation on the monotonicity of the model

To ease the analysis we consider the monotonicity of the pre-synaptic input of the edge detector neuron, $R(t)$ (Fig. 3, D and E), instead of the monotonicity of the neuron membrane potential, $M(t)$ (Fig. 3, F and G). We show that for small enough value of C the total current that is being injected to the edge detector neuron is a decreasing function of P , i.e., $d/dP \int_0^\infty R(t) dt < 0$. Assuming that the derivatives of $U(t, \tau_2)$ with respect to P and τ_2 are continuous, it holds that

$$\frac{d}{dP} \int_0^\infty R(t) dt = - \frac{d}{dP} \int_0^\infty \frac{d}{d\tau_2} \tilde{U}(t, \tau_2) dt = - \int_0^\infty \frac{d^2}{dP d\tau_2} \left[\frac{1 - e^{-U(t, \tau_2)/C}}{1 + e^{-U(t, \tau_2)/C}} \right] dt \quad (\text{B11})$$

Using the derivative chain rule we can further simplify Eq. B11

$$\frac{d}{dP} \int_0^\infty R(t) dt \cong \int_0^\infty - \left(\frac{d^2}{dP d\tau_2} U \right) H dt \quad \text{where } H = \frac{2e^{-U/C}}{C^2(1 + e^{-U/C})^2} \left(C - U \frac{1 - e^{-U/C}}{1 + e^{-U/C}} \right) \quad (\text{B12})$$

Note that for simplifying H we used the approximation for $U(t, \tau_2)$ (Eq. B6), which justifies the following approximation

$$\left(\frac{d}{dP} U \right) \left(\frac{d}{d\tau_2} U \right) \cong \left(\frac{d^2}{dP d\tau_2} U \right) U$$

Since $-d^2U/dP d\tau_2$ is positive and bounded for $0 \leq t < \infty$, U does not depend on C , $\lim_{C \rightarrow 0} H(U) = 0$, and $\lim_{t \rightarrow \infty} H(U) = 0$ we can use the following approximation

$$\int_0^\infty \frac{d^2 U}{dP d\tau_2} H dt \cong \int_0^{t_1} \frac{d^2 U}{dP d\tau_2} H dt \quad \text{such that } U(t_1) = kC$$

where the degree of approximation is determined by the parameter k , independently of C . By using small enough value of C and therefore small enough value of t_1 , we can use the approximations

$$U(t) \cong \frac{AP}{2\tau_2^2 P_0 D^n} t^{n+2} \quad \text{and} \quad \frac{d}{dt} U(t) \cong \frac{(n+2)(AP)^{1/(n+2)}}{(2\tau_2^2 P_0 D^n)^{1/(n+2)}} U^{[1-1/(n+2)]}(t)$$

to transform Eq. B12 as follows

$$\begin{aligned} \int_0^{t_1} -\frac{d^2U(t)}{dPd\tau_2} H[U(t)] dt &= \frac{2}{P\tau_2} \int_0^{t_1} U(t) H[U(t)] dt \\ &= g \int_0^{t_1} U^{1/(n+2)}(t) H[U(t)] \frac{d}{dt} U(t) dt = g \int_0^{kC} U^{1/(n+2)} H(U) dU \quad (B13) \end{aligned}$$

where

$$g = \frac{2(2\tau_2^2 P_0 D^n)^{1/(n+2)}}{P\tau_2(n+2)(AP)^{1/(n+2)}} > 0$$

Since

$$H(U) > 0 \quad 0 < U < u_0 \quad u_0 \cong 1.54C$$

$$H(U) < 0 \quad u_0 < U$$

and

$$\int_0^{u_0} H(U) dU + \int_{u_0}^{\infty} H(U) dU = 0$$

and since $U^{1/(n+2)}$ is monotonically increasing, it holds that

$$\frac{d}{dP} \int_0^{\infty} R(t) dt \cong g \int_0^{kC} U^{1/(n+2)} H(U) dU \cong g \int_0^{\infty} U^{1/(n+2)} H(U) dU < 0 \quad (B14)$$

Analysis of the spike count as a function of the input parameters

Two assumptions are made to analyze the model predictions regarding spike counts, as illustrated in Fig. 3G. First, it is assumed that the neuron fires as long as its membrane potential is above the fixed threshold level. Second, it is assumed that the firing rate is linearly proportional to the level of the membrane potential above the threshold. Formally, let $M(t)$ denote the membrane potential of the edge detector neuron and T denote the fixed threshold level, then the total spike count $S(P, D, n)$ is given by

$$S(P, D, n) = \int_{L_1}^{L_2} [M(t) - T] dt \quad \text{where} \quad M(L_1) = M(L_2) = T \quad (B15)$$

This approach to approximating the spike counts is valid only for $t < D$, since the expression for $M(t)$ (Eq. B8) is valid only for $t < D$ (where D denotes the onset rise time). Therefore, if $L_2 > D$, the approximation of the spike counts is invalid. These limitations restrict further analysis to onsets of sufficiently long duration and large step size.

Under these assumptions, the spike count of the model in response to a power-function onset is a function of P/D^n , since L_1 , L_2 , and $M(t)$ are all functions of P/D^n .

The dependence of the spike count on P/D^n is consistent with experimental (see Fig. B1A) and simulated findings. In particular, since the proof for this relationship is based on the approximation described in Eq. B3, it is expected that the relationship would not hold for parameter values that yield poor approximations. For example, low P values yield poor approximation for these equations and result in spike count that seems to be uncorrelated with P/D^n (e.g., $P \leq 20$ dB, thick lines in Fig. B1A). This fact is the main reason for the failure of the model to fit the data of Bregman et al. (1994b, see Fig. 14).

However, the above analysis does not explain the experimental and simulated relationship between the spike count and the stimulus pressure at the moment of first-spike generation. For the parameter ranges in which the approximations hold, the first-spike latency is a monotonic decreasing function of P/D^n , the stimulus peak pressure at first-spike

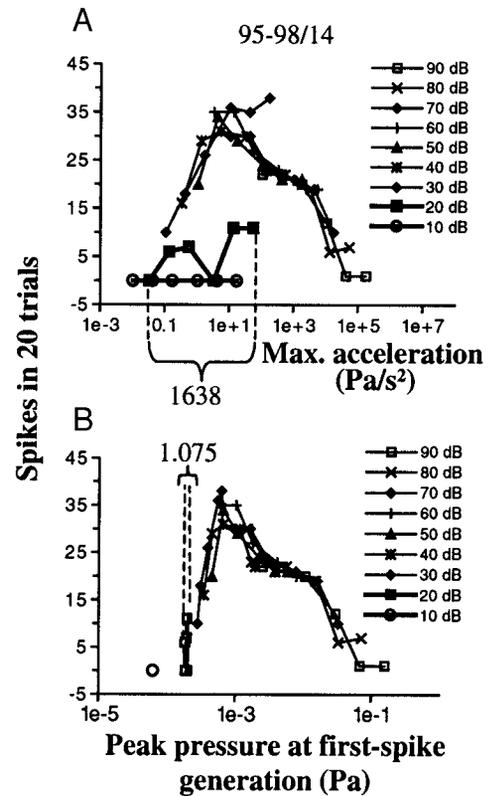


FIG. B1. Spike-count data from Heil (1997b) replotted as a function of maximum acceleration of peak pressure (A) and as a function of stimulus level at the time of 1st-spike generation (B). Note that the data are organized along iso-plateau-level curves and not along iso-rise-time curves. This figure demonstrates that the spike-count data for stimuli above 20 dB SPL converge along a single curve also when presented as a function of maximum acceleration (A). See more details in APPENDIX B.

latency is a decreasing function of P , and therefore the dependence of the spike counts on P/D^n can be transformed into a dependence of the spike counts on the stimulus peak pressure at first-spike latency. To explain why the spike counts are still approximate functions of the stimulus peak pressure at first-spike latency even when the approximations fail, note that the most obvious departures from the approximations occur when P is small. At these lower levels, the spike counts are no longer functions of P/D^n . At these lower levels, the values of P/D^n can vary over order of magnitudes. For example, when $n = 2$ and D covers a range of 4.2:170, P/D^n would cover, for the same P , a range of 1:1,638 (see Fig. B1A). On the other hand, the sound peak pressure at the time of first-spike generation varies much less with D (for example, in Fig. B1B it covers a range of only 1:1.075). Thus plotting spike counts as a function of the sound peak pressure at the time of first-spike generation causes the spike-count curves to better overlap also at these lower values of P (e.g., $P \leq 20$, thick lines in Fig. B1B), but is not an essential feature of the model.

M. Furst helped collect the psychoacoustical data. L. Ahdut, N. Ulanovsky, and G. Jacobson helped collect the electrophysiological data. We thank P. Heil for helpful comments on an earlier version of this manuscript.

This research was supported by a grant from the Human Frontiers Science Program.

Present address of A. Fishbach: Dept. of Biomedical Engineering, Johns Hopkins University, 505 Traylor Bldg., 720 Rutland Ave., Baltimore, MD 21205.

REFERENCES

AGMON-SNIR H AND SEGEV I. Signal delay and input synchronization in passive dendritic structure. *J Neurophysiol* 70: 2066–2085, 1993.

- BARTH CD AND BURKARD R. Effects of noise bursts rise time and level on the human brainstem auditory evoked response. *Audiology* 32: 225–233, 1993.
- BREGMAN AS, AHAD PA, AND KIM J. Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *J Acoust Soc Am* 96: 2694–2703, 1994a.
- BREGMAN AS, AHAD PA, KIM J, AND MELNERICH L. Resetting the pitch-analysis system. 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone. *Percept Psychophys* 56: 155–162, 1994b.
- DRULLMAN R. Temporal envelope and fine structure cues for speech intelligibility. *J Acoust Soc Am* 97: 585–592, 1995.
- DRULLMAN R, FESTEN HM, AND PLOMP R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 95: 1053–1064, 1994a.
- DRULLMAN R, FESTEN HM, AND PLOMP R. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am* 95: 2670–2680, 1994b.
- EGGERMONT JJ. Differential effects of age on click-rate and amplitude-modulation-frequency coding in cat primary auditory cortex. *Hear Res* 65: 175–192, 1993.
- GERSTNER W. Spiking neurons. In: *Pulsed Neural Networks*, edited by Maass W and Bishop CM. Cambridge, MA: MIT Press, 1999a.
- GERSTNER W. Population of spiking neurons. In: *Pulsed Neural Networks*, edited by Maass W and Bishop CM. Cambridge, MA: MIT Press, 1999b.
- HATTORI T AND SUGA N. The inferior colliculus of the mustached bat has the frequency-vs-latency coordinates. *J Comp Physiol [A]* 180: 271–284, 1997.
- HEIL P. Auditory onset responses revisited. I. First-spike timing. *J Neurophysiol* 77: 2616–2641, 1997a.
- HEIL P. Auditory onset responses revisited. II. Response strength. *J Neurophysiol* 77: 2642–2660, 1997b.
- HEIL P. Further observations on the threshold model of latency for auditory neurons. *Behav Brain Res* 95: 233–236, 1998.
- HEIL P AND IRVINE DRF. On determinants of first-spike latency in auditory cortex. *Neuroreport* 7: 3073–3076, 1996.
- HEIL P AND IRVINE DRF. First-spike timing of auditory-nerve fibers and comparison with auditory cortex. *J Neurophysiol* 78: 2438–2454, 1997.
- HEIL P AND IRVINE DRF. Functional specialization in auditory cortex: response to frequency-modulated stimuli in the cat's posterior auditory field. *J Neurophysiol* 79: 3041–3059, 1998a.
- HEIL P AND IRVINE DRF. The posterior field P of cat auditory cortex: coding of envelope transients. *Cereb Cortex* 8: 125–141, 1998b.
- HEIL P, RAJAN R, AND IRVINE DRF. Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hear Res* 76: 188–202, 1994.
- HEWITT MJ AND MEDDIS R. Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse. *J Acoust Soc Am* 87: 1813–1816, 1990.
- HEWITT MJ AND MEDDIS R. A computer-model of amplitude modulation sensitivity of single units in the inferior colliculus. *J Acoust Soc Am* 95: 2145–2159, 1994.
- KITZES LM, GIBSON MM, ROSE JE, AND HIND JE. Initial discharge latency and threshold considerations for some neurons in cochlear nucleus complex of the cat. *J Neurophysiol* 41: 1165–1182, 1978.
- MARR D. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. San-Francisco, CA: Freeman, 1982.
- MCCORMICK DA, CONNORS BW, LIGHTHALL JW, AND PRINCE DA. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J Neurophysiol* 54: 782–806, 1985.
- MENDELSON JR, SCHREINER CE, AND SUTTER KL. Functional topography of cat primary auditory cortex: response latencies. *J Comp Physiol [A]* 181: 615–633, 1997.
- NELKEN I, ROTMAN Y, AND BAR-YOSEF O. Responses of auditory cortex neurons to structural features of natural sounds. *Nature* 397: 154–157, 1999.
- OLSEN JF. Sensitivity of medial geniculate neurons in the squirrel monkey to rate of rise. *Soc Neurosci Abstr* 20: 321, 1994.
- ONISHI S AND DAVIS H. Effects of duration and rise time of tone bursts on evoked V potentials. *J Acoust Soc Am* 44: 582–591, 1968.
- PHILLIPS DP. Effect of tone-pulse rise time on rate-level functions of cat auditory cortex neurons: excitatory and inhibitory processes shaping responses to tone onset. *J Neurophysiol* 59: 1524–1539, 1988.
- PHILLIPS DP. Factors shaping the response latencies of neurons in the cat's auditory cortex. *Behav Brain Res* 93: 33–41, 1998.
- PHILLIPS DP AND BURKARD R. Response magnitude and timing of auditory response initiation in the inferior colliculus of the awake chinchilla. *J Acoust Soc Am* 105: 2731–2737, 1999.
- PHILLIPS DP, SEMPLE MN, AND KITZES LM. Factors shaping the tone level sensitivity of single neurons in posterior field of cat auditory cortex. *J Neurophysiol* 73: 674–686, 1995.
- REES A AND MÖLLER AR. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hear Res* 10: 301–330, 1983.
- RODIECK RW. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res* 5: 583–601, 1965.
- SACHS MB AND ABBAS PJ. Rate versus level functions for auditory nerve fibers in cats: tone burst stimuli. *J Acoust Soc Am* 56: 1835–1847, 1974.
- SCHREINER CE AND LANGNER G. Central coding of temporal patterns. In: *Auditory Function, Neurobiological Bases of Hearing*, edited by Edelman G, Gall W, and Cowan W. New York: Wiley, 1988a.
- SCHREINER CE AND LANGNER G. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60: 1823–1840, 1988b.
- SCHREINER CE AND MENDELSON JR. Functional topography of cat primary auditory cortex: distribution of integrated excitation. *J Neurophysiol* 64: 1442–1459, 1990.
- SHANNON RV, ZENG FG, KAMATH V, WYGONSKI J, AND EKELID M. Speech recognition with primarily temporal cues. *Science* 270: 303–304, 1995.
- SLANEY M. *Auditory Toolbox Version 2*. Interval Research Corporation, 1998. (Tech. Rep. No. 1998-010)
- SMITH RL. Encoding of sound intensity by auditory neurons. In: *Auditory Function, Neurobiological Bases of Hearing*, edited by Edelman G, Gall W, and Cowan W. New York: Wiley, 1988.
- SUGA N. Responses of inferior collicular neurons of bats to tone bursts with different rise times. *J Physiol (Lond)* 217: 159–177, 1971.
- SUTTER ML AND SCHREINER CE. Topography of intensity tuning in cat primary auditory cortex—single-neuron versus multiple-neuron recordings. *J Neurophysiol* 73: 190–204, 1995.
- TURNER CW, RELKIN EM, AND DOUCET J. Psychophysical and physiological forward masking studies: probe duration and rise-time effects. *J Acoust Soc Am* 96: 795–800, 1994.