Level-Dependent Latency Shifts Quantified Through Binaural Processing

Ida Siveke,1 Christian Leibold,1,2 Katharina Kaiser,1 Benedikt Grothe,1,2 and Lutz Wiegrebe1,2

1Division of Neurobiology, Department Biologie II, Ludwig-Maximilians-Universität München; and 2Bernstein Center for Computational Neuroscience Munich, Planegg-Martinsried, Germany

Submitted 28 April 2010; accepted in final form 7 August 2010

Siveke I, Leibold C, Kaiser K, Grothe B, Wiegrebe L. Level-dependent latency shifts quantified through binaural processing. J Neurophysiol 104: 2224–2235, 2010. First published August 11, 2010; doi:10.1152/jn.00392.2010. The mammalian binaural system compares the timing of monaural inputs with microsecond precision. This temporal precision is required for localizing sounds in azimuth. However, temporal features of the monaural inputs, in particular their latencies, highly depend on the overall sound level. In a combined psychophysical, electrophysiological, and modeling approach, we investigate how level-dependent latency shifts of the monaural responses are reflected in the perception and neural representation of interaural time differences. We exploit the sensitivity of the binaural system to the timing of high-frequency stimuli with binaurally incongruent envelopes. Using these novel stimuli, both the perceptually adjusted interaural time differences and the time differences extracted from electrophysiological recordings systematically depend on overall sound pressure level. The perceptual and electrophysiological time differences of the envelopes can be explained in an existing model of temporal integration only if a level-dependent firing threshold is added. Such an adjustment of firing threshold provides a temporally accurate neural code of the temporal structure of a stimulus and its binaural disparities independent of overall sound level.

INTRODUCTION

Precise temporal coding is the hallmark of the auditory system. Like no other sensory modality, the auditory system relies on the neural analysis of spike timing for both object localization and identification. At the same time, the peripheral auditory system has to cope with a huge variability in the loudness of sounds; the dynamic range of natural acoustic input spans at least five orders of magnitude. Several electrophysiological studies have shown that the processing of temporal features highly depends on the overall sound level (Heil 1998; Heil and Irvine 1997; Joris and Yin 1992; Palmer and Shackleton 2009; Tollin and Yin 2005; Wallace et al. 2002). Recent studies on temporal integration of sounds have revealed that neural first-spike latency in the auditory nerve can be explained by the temporal integration of the stimulus pressure envelope (Heil and Neubauer 2003). With increasing sound level, the pressure-envelope amplitude increases, and consequently, a shorter integration time is needed to exceed spiking threshold, i.e., the spiking latency decreases. It is unclear, however, whether this pressure-envelope integration is a sufficient descriptor for ongoing inputs to the binaural system at moderate or high sound levels.

In the binaural system, extraordinary temporal precision is required for the neural coincidence detection that underlies our capability to localize sounds in azimuth (Sandel et al. 1955). Experiments have demonstrated that neural tuning to interaural time differences (ITDs) in the medial superior olive is stable against changes of overall sound level, and this level invariance is further stabilized at higher stages of the auditory system (Cook et al. 2003; Goldberg and Brown 1968; Pecka et al. 2010; Pena et al. 1996). Also psychophysically, it has been shown that sensitivity to ITDs is stable across a wide range of stimulus levels (Blauert 1997; Dreyer and Oxenham 2008). It has been proposed that level-dependent latency shifts may contribute to the processing of interaural intensity differences (IIDs, “latency hypothesis”), (Grothe and Park 2000; Hirsch et al. 1985; Irvine et al. 1995, 2001; Joris and Yin 1995; Park et al. 1996, 1997; Pollak 1988; Yin et al. 1985). However, level-dependent latency shifts should not interfere with binaural analysis as long as both monaural inputs are affected equally. The neural processing underlying this level-invariant encoding of temporal stimulus properties is still poorly understood, in particular because even phase locking, as a basic measure of precise temporal coding, is sensitive to changes in overall sound level in the auditory nerve (Joris et al. 2004). Such level dependence is also seen in the temporal encoding of the envelopes of high-frequency tones (Dreyer and Delgutte 2006).

In this study, we take advantage of the binaural system’s exquisite temporal sensitivity to quantify the temporal integration and the resulting perceptual and spike latency preceding binaural processing in the auditory brain stem. We exploit recent findings that interaural envelope ITDs (eITDs) of high-frequency tones are binaurally analyzed with a precision similar to that of low-frequency tones (Bernstein and Trahiotis 2002). Physiological data obtained from the auditory midbrain support these psychophysical findings (Griffin et al. 2005). Interestingly, the envelopes of high-frequency tones can be manipulated in ways that the fine structure of low-frequency tones cannot: at high frequencies, stimuli can be constructed that carry different binaural temporal properties although they occupy the same frequency region in the two ears. We manipulated the envelope properties of high-frequency stimuli by changing the rise and fall times of the envelopes and presented these stimuli at different sound levels. The experimental paradigm is illustrated in Fig. 1, which shows qualitatively how changes in the rise time and in the sound level of tone pips lead to shifts in latency. With increasing sound level, the latency difference induced by the difference in rise time converges to zero (cf. Fig. 1C). Using these new stimuli in a binaural experimental paradigm, we have directly quantified level-dependent latency shifts both psychophysically and electrophysiologically. The psychophysical and electrophysiological data show that even at the highest sound levels tested, the rise-time differences between the pips in the left and right ear are reflected in the eITDs. This result is inconsistent with current models of neural latency, even when cochlear compressive nonlinearity is taken into account. This limitation can only be overcome by a level-dependent
adjustment of firing threshold. Thus in line with previous modelling- and in vitro studies in birds (Cook et al. 2003; Dasika et al. 2005; Kuba et al. 2002; Kuznetsova et al. 2008), the current results indicate the importance of neural adaptation preceding binaural processing. Preliminary data of this study were presented at the 15th International Symposium in Hearing, Salamanca, Spain.

METHODS

Rationale

The stimuli in this study were trains of tone pips. The pips consist of a high-frequency pure-tone carrier modulated with temporally asymmetric envelopes. If such a pip train is presented binaurally but temporally reversed in one ear, as illustrated in Fig. 1, it cannot be physically matched by simple interaural time shifts. Nevertheless, the binaural system can produce a reliable perceptual match that in turn reflects the way the stimulus envelopes are processed preceding binaural analysis. Note that although no physical match can be produced, we refer to the result of the perceptual match as the adjusted eITD. If the system is forced to produce a binaural match, one would expect that the eITD that produces a centralized binaural image does depend on overall sound level: if the sound level is low (solid lines in Fig. 1A), one would expect that the pips in the ear with the steeper rise time (black lines) have to be strongly delayed to compensate for the rise-time difference between the ears (large dT in Fig. 1B). If the sound level is high (dotted lines in Fig. 1A), one would expect that the overall latency decreases because the envelope integral exceeds the firing threshold earlier (B), but again the pips in the ear with the steeper rise time (black lines) have to be somewhat delayed to compensate for the rise-time difference between the ears (small dT in Fig. 1B). Thus with increasing sound level, not only the overall latency is expected to decrease, but also the latency difference elicited by the difference in rise time (Fig. 1C). Current models of neural latency predict that this latency difference approximates zero at high sound levels irrespective of the type of cochlear compressive nonlinearity.

Psychophysics

STIMULI. For the psychophysics, the stimuli were trains of eight tone pips with a carrier frequency of 5 kHz. The interval between the pips was fixed at 10 ms. The pips consisted of a linear rise and decay without steady state. For the left-ear stimuli, the rise time was 1, 2, 4, or 8 ms, and the decay time was always double the rise time, resulting in tone pips with a duration of 3, 6, 12, or 24 ms. The pip trains in the right ear were either identical to the left-ear trains or they were temporally reversed. To preclude residual carrier-periodicity related cues, the pure-tone carrier was refreshed for each ear and each stimulus with randomized phase. Experiments were run with a continuous background noise (8 dB SPL/Hz) low-pass filtered at 1,500 Hz to preclude the binaural analysis of low-frequency aural distortion products. Stimuli were generated in MATLAB (The Math Works, Natick, MA) at a sampling rate of 48 kHz and played via an RME-Audio DIGI 96/8 PST soundcard (Synthax, Haimhausen, Germany) into AKG K240 DF circumaural headphones (Vienna). The earphones were digitally equalized in magnitude and phase based on its binaural impulse responses measured on a Bruel and Kjaer 4153 artificial ear (Naerum, Denmark). The listeners were seated in a double-walled sound-attenuating booth (G+H Schallschutz, Ludwigshafen, Germany).

PROCEDURE. In a two-alternative, forced-choice paradigm without feedback, listeners were asked to judge the lateralization of a test stimulus compared with a reference stimulus. The reference stimulus was always presented first. It consisted of a diotic pip train with identical rise and decay times in both ears and no eITD. The test stimulus followed the reference after a 500-ms pause and was presented with a variable eITD. For the first trial of each run, this variable eITD was set randomly in a range within ±500 μs. In each trial, listeners judged whether the test stimulus was lateralized left or right of the reference stimulus. In an adaptive procedure, the eITD of the test stimulus was changed to match the lateralization of the reference stimulus. For the 1st and 2nd reversal (a change in lateralization direction), eITD was changed in steps of 160 μs, for the 3rd to 5th reversal, the step size was reduced to 80 μs, and for reversals 6–11, it was reduced to 40 μs. The adjusted eITD in a given experimental run is given as the mean eITD across reversals 6–11. Individual data are based on at least six runs per experimental condition. An experimental session consisted of three standard conditions and three test conditions in randomized order: in the standard condition, the test stimulus consisted also of a pip train with identical rise and decay times in the two ears. In the test condition, the pip train in the right ear was temporally reversed compared with the left ear, i.e., while pips in the left ear had a steep rise and shallow decay time, pips in the right ear had a shallow rise and a steep decay time.

Lateralization was measured as a function of the pip-train sensation level. For each listener, the absolute threshold for the reference stimulus was measured in an adaptive two-alternative, forced-choice experiment with feedback following a three-down, one-up paradigm. The level step size was 10, 5, and 2 dB for reversals 1–2, 3–5, and 6–11, respectively.
Pip-train levels for the main experiment were based on the individual absolute thresholds for each listener and pip duration.

LISTENERS. At least four normal hearing listeners, aged between 21 and 35 of both genders took part in each experiment. Listeners for a pip duration of 6 ms were different from those that generated the data for the different pip durations.

ELECTROPHYSIOLOGY. Single-unit responses were recorded from two nuclei in the gerbil auditory brain stem, the lateral superior olive (LSO) and the dorsal nucleus of the lateral lemniscus (DNLL). Overall we chose to record single-neuron responses from the output stage of the nuclei of interest, i.e., monaural responses from the LSO as an output stage of the ventral cochlear nucleus and from the DNLL as an output stage of the medial superior olive (MSO). Responses were collected from high-frequency units that are insensitive to the carrier phase but that lock onto the envelope period of the presented pip trains. The LSO recordings were performed to assess the monaural inputs, not binaural processing, capitalizing on the fact that cochlear-nucleus spherical bushy cells project both to the LSO and MSO and that the LSO faithfully preserves the bushy-cell firing characteristics (cf. Hassfurther et al. 2009; Magnusson et al. 2008). Consequently stimulation in the LSO was monaural from the ipsilateral side. In fact most LSO units did not respond to binaural pip-train stimulation due to the inhibitory effect of the contralateral ear.

Assessment of binaural electrophysiological representation of the pip-train stimuli is again based on the neural responses of the output stage, in this case the DNLL.

ANIMALS. Recordings were obtained from 18 adult Mongolian gerbils (Meriones unguiculatus). They were of both sexes, 2–5 mo of age, and weighting between 50 and 90 g. Single cells were recorded in two different brain stem nuclei, the LSO (n = 32), and the DNLL (n = 50). All experiments were approved according to the German Tierschutzgesetz (AZ 55.2-1-54-2531-57-05). The detailed methods in terms of surgical preparation, acoustic stimulus delivery, stimulus calibration, and recording techniques have been described previously (Siveke et al. 2006).

SURGERY. The animals were anesthetized by a physiological NaCl solution containing ketamine (20%) and xylacine (2%). The adequate state of the anesthetized animal was ascertained regularly throughout the experiment by testing the absence of motor reflexes and monitoring breathing rate. To secure a constant state of anesthesia, supplementary doses of the same mixture were given subcutaneously with a continuous rate of 1.7 μl/min or, in some experiments, 0.05–0.1 ml every 30 min. Constant body temperature (37–39°C) was maintained using a thermostatically controlled heating blanket. During recordings, the animal was placed in a sound-attenuated chamber and mounted in a custom-made stereotactic instrument allowing reproducible positioning of the skull (Schuller et al. 1986). Ear molds were attached to the head, allowing the insertion and fixation of earphones and probe-tube microphones. To gain access to the LSO, a craniotomy was performed lateral to midline of the skull and caudal to the posterior aspect of the transverse sinus. The underlying cerebellum was partially aspirated to expose the floor of the fourth ventricle. To access the DNLL, a small craniotomy was performed lateral to the midline and caudal to the interaural axis. To access the LSO, electrode penetrations (tilted 20° rostrally) were performed 1.4–1.8 mm lateral to the midline and 4.3–4.7 mm caudal of the interaural axis, to access the DNLL, electrode penetrations (tilted 10 or 5° laterally) were performed 1.3–2.0 mm lateral to the midline and 0.5–0.8 mm caudal of the interaural axis.

RECORDING PROCEDURE AND GENERAL NEURAL CHARACTERIZATION. Single-cell responses were recorded extracellularly using glass electrodes filled with 1 M NaCl (~10 MΩ) or 2% horseradish peroxidase (HRP, Sigma-Aldrich) diluted in 10% NaCl. The recording electrode was advanced under remote control using a motorized micromanipulator (Digimatic, Mitutoyo, Neuss, Germany) and a pipette (Inchworm Controller 8200, EXFO Burleigh Products Group). The amplified (Toelner 7607, Herdecke) and filtered (VBF/3, Kemo, Beckenham, United Kingdom) action potentials were fed into the computer via an A/D converter (RP2-1, TDT-Tucker Davis Technologies). Clear isolation of action potentials from a single neuron (signal to noise ratio >5; see waveform of the recorded spikes in the insets of Figs. 4B, 5B, and 7B) was guaranteed by visual inspection (stable size and shape) on a spike-triggered oscilloscope and by off-line spike cluster analysis (Brainware, Jan Schnupp, TDT). The recording sites of 24 of the 32 LSO neurons that we analyzed were marked with HRP that was ejected through the recording pipette via a current source (HV-TR 150; NPI, Tamm, Germany) using a current of 1 μA that was applied to the pipette for 1 min. DNLL recordings were made in locations that were verified to be within the DNLL in previous studies (Siveke et al. 2006). In three DNLL experiments, the last electrode position was marked by a current-induced lesion (5 mA for 5 s after a lethal injection of barbital had been applied) using metal electrodes (5 MΩ). Typical recording periods lasted 10–14 h, after which animals were injected with a lethal dose of barbital. The animals in which a HRP injection or a current-induced lesion was performed were perfused intracardially with heparinized physiological NaCl solution followed by 4% paraformaldehyde solution. The brains were sectioned frontally and sections stained for HRP with diaminobenzidine, intensified by cytochrome and counterstained by standard protocol using Neutral Red. Sections of the current-induced lesions were Nissl-stained. The recording sites were verified using standard light microscopy.

Stimuli were generated at 48 kHz sampling rate in Matlab, converted to analog signals (RP2-1, TDT), attenuated (PA5, TDT) and delivered to the ear-phones (Sony MDR-EX70 LP, Tokyo, Japan). For general neural characterizations, stimulus duration was 200 ms plus 5-ms raised-cosine rise and fall times with a repetition period of 500 ms. As search stimulus, we used uncorrelated binaural noise bursts with an IID of 0 dB. Using pure tones, we first determined audiovisually the neuron’s characteristic frequency (CF) as that frequency that elicited a response at the lowest intensity, neuronal threshold (thr). For all neurons, both CF and the corresponding thr were later confirmed by a careful off-line analysis of the frequency versus level response areas (9 frequencies, step size CF/5, 10 dB steps) and rate level function at CF (8 dB steps). In the LSO, sensitivity to IIDs was assessed by presenting binaural tones at CF (ITD = 0) with a combination of different IIDs (thr-30 to thr +20 dB, 10 dB steps) holding the excitatory (ipsilateral) ear constant at 20 dB above thr. Neurons were defined as IID sensitive if contralateral (inhibitory) stimulation reduced the maximal response elicited by ipsilateral (excitatory) stimulation by >50%. In the DNLL, sensitivity to eITDs was assessed by measuring noise delay functions (NDFs) presented with different ITDs over a range of ±3.5 ms (step size: 125 μs) determined from 2 to 16 repetitions. A unit was considered as eITD sensitive if the noise-delay function was modulated by ±50% (i.e., if the minimum discharge rate was less than half of the maximum rate). The ITD eliciting maximal spike rate was defined as the neuron’s best ITD.

PIPTRAIN STIMULATION. The pips consisted of the same envelopes as in the psychophysical experiments. The carrier frequency was set to the cell’s CF. The pips were presented in a 4-s train at a repetition rate of 40 Hz. Thus one train consisted of 160 pips. For the LSO, five repetitions of the pip trains and temporally reversed pip trains were presented monaurally on the ipsilateral ear in a randomized sequence. Pip trains were presented with different pip durations (3, 6, 12, and 24 ms) and different sound levels (18 to 50 dB above threshold in 8 dB steps).

For the DNLL, the pip trains were presented binaurally (see Fig. 8B, inset). In the contralateral ear, the temporal structure of the envelope was constant; in the ipsilateral ear, the stimulation was identical to the LSO.

J Neurophysiol • VOL 104 • OCTOBER 2010 • WWW.JN.ORG

Downloaded from http://jn.physiology.org/ on June 11, 2017 by 10.220.33.1 on June 11, 2017
monaural stimulation, i.e., the temporal envelope of the pip train was equal to the contralateral side or temporally reversed. Furthermore in the ipsilateral ear, the interpip interval was used to generate time-variant eITDs: within 1 s of pip-train stimulation, the delay of the ipsilateral pip was varied across a range spanning ± one-quarter of the pip duration. This delay range was centered on the best ITD of the neuron. For a 40 Hz pip rate, the eITD range thus consisted of 40 different eITDs. As for the LSO stimulation, pip trains were presented at different pip durations and different sound levels above threshold.

**ANALYSIS.** For all neurons recorded in the LSO and the DNLL, we first analyzed phase locking, quantified in terms of vector strength (Goldberg and Brown 1968) to the period of the pure-tone carrier and the envelope of the pips. Only neurons with significant phase locking to the envelope of the pips (see Figs. 5D and 8B) but not to the carrier (P < 0.01; Rayleigh test) were used for further analysis. To calculate the time shift in the period histograms recorded with the standard and time-reversed pips in the LSO, we performed a cross-correlation between the two histograms. A Gaussian was fitted to the cross-correlation function (Fig. 8C) and the lag of the maximum of the fit was taken as the time shift. The binaural DNLL response rates were plotted as a function of the eITD and fitted with a Gaussian for both standard and ipsilaterally reversed pip trains. The differences between the eITDs at which the two Gaussians had their maxima were taken as the eITD change. From the population of recorded neurons (n = 50), only the neuronal data with a correlation coefficient > 0.7 between the data and the Gaussian fits were used for further analysis (n = 30).

**Model fitting**

Two variants of the threshold integral model proposed by Heil and Neubauer (2003) were employed to fit the psychophysical and electrophysiological data. The model results were based on predicting the latencies \( L_l \) and \( L_r \) of the neuronal activities evoked by the acoustic stimuli at the left and right ear, respectively. The matched eITD was then derived from the binaural latency difference \( L_r - L_l \). The threshold integrals \( S(L) \) that were used for the latency prediction are described and explained in RESULTS. The parameters of the models were obtained from a numerical fitting procedure. Using the MATLAB routine “fminsearch,” we minimized the root mean squared error

\[
E = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\tau_i - \Delta_t)^2}
\]

between all measured eITD changes \( \tau \) (all pip durations and sound-pressure levels) and the respective model predictions \( \Delta_t \). The fitting procedure was iterated for different starting values until the fit parameters converged. To have a better comparison between different experimental conditions, we also give the relative error, rel. err. = \( E^2 / \text{var}(\tau) \), which is the squared error normalized by the variance of the data.

**RESULTS**

The eITDs adjusted by the listeners to centralize the incongruent pip trains are shown in Fig. 2. Data are averaged across listeners; the error bars represent standard errors. At a sensation level of 20 dB, the listeners adjusted the temporally reversed 6-ms pip train in the right ear (with a 4-ms rise and 2-ms decay time) to start ~1,150 µs earlier than the pip train in the left ear (with a 2-ms rise and 4-ms decay time). With this eITD, the perceived lateralization matched that of a diotic 6-ms pip train with 2-ms rise time and 4-ms decay time. With increasing sensation level, the adjusted eITD decreased monotonically, but, even at a high sensation level of 60 dB, the listeners still required an eITD of ~580 µs to compensate for the different rise times in the two ears. Note that this experimental finding is in contrast to the expectation outlined in Fig. 1 and the corresponding text: based on this logic, one would expect that at a high sensation level, the rise-time difference is irrelevant. This would result in an adjusted eITD near 0 µs, i.e., the incongruent pip trains should be perceived close to the center without an additional delay of the pip train with the steeper rise time.

Not surprisingly, the adjusted eITD systematically depended on the overall duration of the rise and decay times. The adjusted eITD increases with increasing overall pip duration for all sensation levels. However, in all conditions, even at the highest tested sensation levels, the adjusted eITD remained > 500 µs. Especially for the short tone-pip durations, the data clearly show a floor effect, i.e., further increases in level do not lead to further decreases in adjusted eITD.

To study whether the eITD shifts can be explained by the level-dependent latency difference \( L_r - L_l \), between the left and right ear evoked neuronal activity, we employed two model variants of the pressure integrator proposed in Heil and Neubauer (2003). In contrast to the original model, in which latencies are predicted based on the sound pressure wave, we assume integration of the pressure envelope \( e(t) \) of the sound wave. Then the binaural latencies \( L_l \) and \( L_r \) of action potentials are assumed to be determined by the threshold crossings of the respective integrated envelopes \( S \) (Fig. 3, A and B). Mathematically this maps to the implicit equation

\[
\Theta = S(L_l) = \int_0^{L_l} \text{d}[e_{\text{a}}(t)]^k
\]

in which we allowed an additional exponent \( k \) as a fit parameter. In the first model variant (called the “latency model,” Fig. 3A), the threshold \( \Theta \) is a constant and independent of sound level. This model thus has two fit parameters, \( k \) and \( \Theta \). In the second model variant (called “latency model with level-dependent firing threshold,” Fig. 3B), the threshold is a function of the signal level. Specifically we modeled the threshold as

\[
\Theta = \Theta_0 + \Theta_1 \alpha
\]

where \( \alpha \) denotes the amplitude of the envelope \( e(t) \) in units of Pascal. The second model thus has three fit parameters, \( k, \Theta_0, \) and \( \Theta_1 \).

When comparing the best model fits to the psychophysically adjusted eITDs, we observed that the latency model with a constant firing threshold cannot reproduce the large eITDs at high sound levels, although the model qualitatively fits the result that in longer pips produce larger eITDs (Fig. 3C). In contrast, the latency model variant that includes a level-dependent threshold can account for the saturation of adjusted eITDs at high sound levels (Fig. 3D). The best fit parameters of the first model variant were \( k = 1.21, \Theta = 0.0033 \) (fitting error: \( E = 415 \mu s, \text{rel. err.} = 2.8 \)). The best fit parameters of...
the second model variant were $k = 1.46$, $\Theta_o = -9.4 \times 10^{-6}$ and $\Theta_i = 0.0697$ Pa (fitting error: $E = 149$ $\mu$s, rel. err. = 0.36).

To measure the monaural inputs to the binaural system in the SOC, neural responses to ipsilateral pip trains were obtained from the LSO of the gerbil. We recorded from 32 cells with CFs between 2 and 13 kHz (mean: 4.7 kHz). Histological verifications of the recording sites are shown in Fig. 4A.

FIG. 4. Histological verification of a recording site and typical auditory responses in the lateral superior olive (LSO). A: Nissl staining of the superior olivary complex (top), schematic drawing of the main nuclei (bottom). The lateral part is on the left and the medial on the right side. On the right, parts of the medial nucleus of the trapezoid body (MNTB) are visible. The typical band of aligned neurons of the medial superior olive (MSO) is in the middle of the picture, the U-shaped LSO with the horseradish-peroxidase (HRP) injection site in the middle of the nucleus is on the left side. B: responses of 6 typical neurons in the LSO to IIDs. Insets: the shapes of the extracellularly recorded action potentials of 2 single neurons. Neurons in the LSO are typically excited from the ipsilateral and inhibited for the contralateral side. Positive IIDs indicate a higher sound pressure level at the ipsilateral side and evoke strong neuronal responses; negative IIDs indicate a higher sound pressure level at the contralateral side and evoke low or no responses.
response (Fig. 5A) with no phase-locking to the carrier but stable phase locking to the envelope period (Fig. 5A, B, and D). The period histograms in response to 6-ms pips (Fig. 5D) show that the spiking latency depends on the direction of the asymmetric envelope in that the cell fires earlier when the pip has a 2-ms rise and a 4-ms decay (Fig. 5, C and D, black graphs). With increasing stimulation level, the response rate increases but the phase difference between the period histograms of the two envelope conditions decreases. This is quantified in the cross-correlograms of the period histograms in Fig. 5E. The peaks in the cross-correlograms shifts from 810 \( \mu s \) at a sound level of 26 dB above threshold to 357 \( \mu s \) at 50 dB above threshold. This time shift is shown for the four different pip durations in Fig. 6A as a function of the sound level above threshold. Qualitatively similar to the psychophysically determined eITDs, the time shift increases with overall pip duration. For each pip duration, the time shift decreases with increasing sound level. Note that even at the highest sound level, the time shifts in the period histograms are still prominent. These main features are preserved in the LSO population average, shown with filled symbols in Fig. 6B. The error bars increase systematically with increasing pip duration. This may result from a less precise neural encoding of the much shallower envelopes of the longer-duration pips.

Electrophysiological results from the gerbil LSO are in good agreement with the eITDs adjusted in the human psychophysical experiments, especially for the longer pip durations of 12 and 24 ms. For shorter pip durations, the LSO data predict smaller eITD changes than observed psychophysically. The latency model with constant threshold (Heil and Neubauer 2003) is unable to predict the LSO data in that it cannot predict the relatively large time shifts seen at high sound levels. For the latency model with a level-dependent firing threshold, the fits to the LSO data capture the main trends but underestimate the time shifts measured at high levels for long pip durations. This is also reflected by the fitting errors \( E = 607 \mu s \) (rel. err. = 0.28) for the latency model and \( E = 533 \mu s \) (rel. err. = 0.22) for the latency model with level-dependent firing threshold.

Recordings in the DNLL were obtained to get an estimate of how the gerbil’s binaural system evaluates the time shifts in the peripheral representation of the pip trains. Again histological verifications of the recording sites are shown in Fig. 7A. We systematically searched for
using the same color code. Dotted lines show the psychophysical data (from Fig. 2) as solid lines. The latency model with threshold adjustment fits to these data. Examples of the noise-delay functions of such units showed sensitivity to eITDs in response to interaurally correlated high-CF binaural units (between 2 and 17 kHz; mean: 5.1 kHz), which were excited also with ipsilateral stimulation (E/E), 8 cells responded to the LSO neurons for the 24-ms pip (black dots; n = 29), and the 3-ms pip (light gray dots; n = 29). The error bars indicate SE. The latency model with threshold adjustment fits to these data are shown as dotted lines; the psychophysical data (from Fig. 2) are shown as solid lines using the same color code.

FIG. 6. Dependence of LSO time shifts on overall sound pressure level and pip duration. A: the time shift of the period histograms of a single LSO neuron (CF = 13 kHz; thr = 20 dB SPL) are plotted against the SPL for a pip duration ranging from 24 ms (black) to 3 ms (light gray). B: population responses (n = 32) are shown as the average time shift of the LSO neurons for the 24-ms pip (black dots; n = 24), the 12-ms pip (dark gray dots; n = 29) the 6-ms pip (gray dots; n = 30), and the 3-ms pip (light gray dots; n = 29). The error bars indicate SE. The latency-model with threshold adjustment fits to these data are shown as dotted lines; the psychophysical data (from Fig. 2) are shown as solid lines using the same color code.

The observed time differences of the binaural LSO responses are much smaller than the theoretically extracted time shifts for the monaural LSO data and also much smaller than observed psychophysically. However, considering the different methods and experimental requirements the three datasets are in good qualitative agreement.

**DISCUSSION**

We have presented a binaurally asymmetric, incongruent stimulus paradigm to quantify the level dependence of both perceptual and neural auditory response latencies. The paradigm exploits two key features of temporal auditory processing: the precise encoding of the envelopes of high-frequency carriers and the exquisite precision of eITD analysis. The used pip stimuli are composed of linear rise and fall times that constitute the envelope for a high-frequency sinusoidal carrier. The stimuli were investigated with respect to human binaural perception as well as their neuronal representations that are revealed by electrophysiological recordings from two different nuclei, the LSO and DNLL, of anesthetized gerbils. These three sets of experimental data are in qualitative agreement: an eITD induced by the binaural asymmetry of the envelopes decreases with decreasing rise time of the envelopes and with increasing sound level. Furthermore all three different experimental datasets showed that the eITDs never decrease to zero but rather asymptote toward a particular eITD for the shortest rise times and most intense stimuli.

To check whether the observed eITDs can be explained by a current model of first-spike latency, we employed two variants of a spike-latency model by Heil and Neubauer (Heil and Neubauer 2003) although the overall fitting errors are rather similar in both models (latency model: E = 153 μs, latency model with level-dependent threshold: E = 150 μs, rel. err. = 0.10 in both cases). In general, the observed time differences of the binaural LSO responses are much smaller than the theoretically extracted time shifts for the monaural LSO data and also much smaller than observed psychophysically. However, considering the different methods and experimental requirements the three datasets are in good qualitative agreement.
Heil and Neubauer proposed a neural correlate of the pressure-envelope integration preceding spike generation to occur at the innerhair cell, auditory-nerve synapse. Recent work has addressed the physiological basis of this pressure-envelope integration. In a constructive discourse, it was shown that the interplay between the stochasticity of synaptic events and a short time constant of synaptic transmission can explain the apparently long time constant of temporal integration at threshold (Krishna 2006; Meddis 2006a,b). In contrast to the modeling work by Heil and Neubauer, Meddis’ (2006a) model also incorporated compressive nonlinear processing in the inner ear. It is important to note that the threshold adjustment cannot be replaced by a simple static compression: such a compressive nonlinearity is often implemented by taking a power law of the pressure envelope (Oxenham and Moore 1994). In the latency model, this power law would simply be compensated for by the fitting parameter k. Thus a static power-law compression alone won’t show the experimentally observed effect on the binaural latency differences if it is not accompanied by a level-dependent threshold. More complex, dynamic compressive mechanisms (e.g., Dau et al. 1996; Tan and Carney 2003) may contribute to the hypothesized threshold adjustment.

LEVEL-DEPENDENCE OF TEMPORAL CODING IN THE AUDITORY NERVE. Dreyer and Delgutte (2006) showed that at the level of the cat auditory nerve, the temporal representation of transposed tones, with an envelope modulation comparable to the current pip trains, deteriorates with increasing sound level. In contrast, the current data show that at the level of the gerbil LSO, the temporal representation of the pip-train envelopes does not deteriorate with increasing sound level. The current data are thus also in agreement with the psychophysical data showing that the sensitivity to eITDs elicited by transposed tones is stable over a wide range of sound levels (Dreyer and Oxenham 2008). These findings also argue for a refinement of temporal envelope encoding at the level of the cochlear nucleus, similar to the refinement of phase locking to pure tones (Joris et al. 2004).
Palmer and Shackleton (2009) have shown that the phase of phase-locked auditory-nerve responses to low-frequency pure tones varies systematically with sound level. A comparable effect is reflected in our data (cf. period histograms in Fig. 5B) obtained with the envelopes of high-frequency tones. While these phase shifts are consistent with a decrease in neural latency with increasing sound level, the level-dependent phase shifts as such do not contribute to the level-dependent eITDs reported here. We always used the same level on the two ears, and thus the phase shifts would apply equally to both ears and cancel out for the eITD estimation.

POSSIBLE NEURAL BASIS FOR THE LEVEL-DEPENDENT THRESHOLD ADJUSTMENT. The physiological basis of the level-dependent threshold proposed in this paper cannot be assessed in the framework of the present study. Our data reveal that the level-dependent threshold adjustment is already present in the input to the SOC. The success of the published model by Heil and Neubauer (2003) to explain the level dependence of auditory-nerve first-spike latency points toward the cochlear nucleus as a neural stage of firing-threshold adjustment.

Experimental evidence for the adjustment of spiking to overall sound level has been seen in several previous studies: Gibson et al. (1985) showed that at the level of the cochlear nucleus, the level of simultaneous background noise strongly influences pure-tone rate-level functions. These authors contributed the observed effects to “two-tone suppression,” i.e., the effects of inhibitory sidebands. Dean et al. (2005) recently showed that comparable effects can be elicited not by frequency but temporal interactions. Specifically, they showed that rate-level functions of auditory-midbrain neurons can adapt to the statistics of sound-level fluctuations. In a follow-up study, the authors showed that this adaptation can be described with time constants of a few hundred milliseconds (Dean et al. 2008). The fact that in the current psychophysical data, adaptation to overall sound level is evident with pip-train stimulation over little more than 100 ms suggests that similar processes may underlie our electrophysiological data. Note, however, that the current data do not allow speculation over the time constant for the threshold adjustment because both the psychophysical and electrophysiological experiments maintained the same sound level over several minutes of stimulation, which precludes the analysis of an adjustment time constant.

INCONSISTENCY OF LSO AND DNLL DATA. The level-dependent eITD changes extracted from the monaural electrophysiological data recorded in the gerbil LSO are larger than the time shifts recorded in the gerbil DNLL with binaural stimulation. A possible reason for this discrepancy may lie in the different stimulation paradigms: the monaural data were recorded with static pip trains while for the binaural data, time-variant inter-pip intervals, creating a binaural beat, were recruited. This was necessary to accommodate the stimulation proto-

![Figure 8](http://jn.physiology.org/)

**FIG. 8.** Response of a DNLL neuron (CF = 3.6 kHz; thr = 25 dB SPL) to the binaural pip train stimuli. A: the raster plot shows a defined acoustical response to a particular envelope eITD of the 6-ms pip train stimuli, which is, according to the “1 Hz beat” stimulation, repeated every second. B: period histograms for the standard (black) and the time reversed (gray) 6-ms binaural pips are plotted for 5 different sound pressure levels ranging from 18 to 50 dB above neuronal threshold. The time base for these period histograms is the pip repetition period, 25 ms. Insets: the schematic diagram of the presented binaural standard and the time-reversed pips. C: the tuning of the neuron to a particular envelope eITD is revealed by the period histograms constructed with the envelope beat period of 1 s. Within each second of the envelope period, the eITD changes linearly and the abscissa is relabeled as eITD axis. The eITD change (Δ) between the envelope period histograms obtained from the standard (black) and the time reversed pip stimulation is obtained by the difference between the shifts of the 2 histogram maxima obtained from Gaussian fits (dotted lines).
data are show as dotted lines.

... bars indicate SE. The latency-model (with threshold adjustment) fits to these...

... with threshold adjustment to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold to overall sound level, can include an adjustment of firing threshold...

... the model was not fitted to the data, the model predictions capture the general level dependence of the adjusted eITDs reasonably well.

FIG. 9. Dependence of DNLL eITD changes on overall sound pressure level and pip duration. A: the eITD changes of the envelope period histograms of a single DNLL neuron (CF = 3.8 kHz; thr = 20 dB SPL) are plotted against the SPL for pip durations ranging from 12 ms (black) to 3 ms (light gray). B: population responses (n = 30) are shown as the average eITD change of the DNLL neurons for the 12-ms pips (black dots; n = 11/22) the 6-ms pips (dark gray dots; n = 13/20) and the 3-ms pips (light gray dots; n = 15/17). The error bars indicate SE. The latency-model (with threshold adjustment) fits to these data as show as dotted lines.

... within a realistic recording time. Although the beat frequency was only 1 Hz, it is possible that the resulting motion cue may influence the eITD sensitivity. Hysteresis effects related to such motion cues were shown for binaural beats elicited with low-frequency pure tones in the mammalian IC (McAlpine et al. 2000; Spitzer and Semple 1993). Another possible reason is that the population characteristics of neurons are different in the DNLL and the LSO: The CFs of the two populations are similar, but their basic binaural properties were different: The LSO neurons were excited from the ipsilateral and inhibited from the contralateral side, where the majority of the DNLL neurons were binaurally excited, similar to neurons in the MSO. It has been shown that bushy cells in the anteroventral cochlear nucleus send collateral projections to the lateral and medial superior olive (Smith et al. 1993; Stotler 1953). Therefore we chose to record from ipsilaterally stimulated LSO cells, which are much easier to record from, assuming that the excitatory input to these cells is similar to the excitatory input to the MSO. Finally DNLL data may deviate from the LSO data because the DNLL is not a principal stage of binaural envelope processing: There are multiple ascending binaural projections bypassing the DNLL in which the envelopes of high-frequency sounds are accurately encoded (Oliver 2000; Pollak et al. 2003). Furthermore persistent inhibition as an important feature of DNLL sounds are accurately encoded (Oliver 2000; Pollak et al. 2003).

... the current paradigm with IID extraction opens new opportunities to study the dynamics of binaural processing.

RELATIONSHIP TO THE LATENCY HYPOTHESIS. Several studies have explored the hypothesis that high-frequency IIDs may be encoded not with interaural differences in spike rate but differences in spike latency (Grothe and Park 2000; Hirsch et al. 1985; Irvine et al. 1995; Joris and Yin 1995; Pollak 1988; Yin et al. 1985; Park et al. 1996, 1997). Irvine et al. (2001) showed that a relatively small proportion of neurons in the LSO encoded high-frequency transient IIDs in terms of latency not rate differences. However, in the LSO of awake bats, this proportion was much higher (Park et al. 1996). In line with previous data, the current LSO data (cf. Fig. 5) show a decrease of spike latency with increasing sound level with a trading ratio of ~77 μs/db. However, as outlined in the preceding text, the current stimuli create short-term time variant IIDs only if a short time constant of IID extraction is assumed. Based on the results of Tollin and Yin (2005), such a short time constant of IID extraction appears plausible. If these short-term IIDs were converted to changes in response latency, the resulting latency shifts would interfere with the extraction of eITDs. The additional latency shifts resulting from the interaural difference in rise time are superimposed on the level-dependent latency shifts. Using the current stimulus paradigm, we cannot disambiguate these two types of latency shifts.

In summary our experimental paradigm provides insights into the level dependence of neural auditory latencies. Recruiting the temporal precision of the high-frequency binaural system, the paradigm allows, for the first time, to quantify the level dependence of auditory neural latencies perceptually. The psychophysical and electrophysiological data show that a modified version of a spike-latency model, which includes an adjustment of firing threshold to overall sound level, can account for the level-dependent perception and peripheral neural encoding of transient sounds.

FIG. 10. Simulation results based on the Zilany et al. (2009) auditory-nerve model (···) compared with the current psychophysical data (—). Although the model was not fitted to the data, the model predictions capture the general level dependence of the adjusted eITDs reasonably well.
Dreyer A, Delgutte B.
Dean I, Harper NS, McAlpine D.
Dean I, Robinson BL, Harper NS, McAlpine D.
Heil P, Neubauer H.
Heil P, Irvine DR.
Griffin SJ, Bernstein LR, Ingham NJ, McAlpine D.
Dreyer AA, Oxenham AJ.
Cook DL, Schwindt PC, Grande LA, Spain WJ.
Blauert J.


