Temporal effects in interaural and sequential level difference perception[a)]

Bernhard Laback[b)]
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, Vienna, A-1040, Austria

Mathias Dietz[c)]
Medizinische Physik, Universität Oldenburg, Carl-von-Ossietzky Straße 9-11, Oldenburg, 26111, Germany

Philip Joris
Laboratory of Auditory Neurophysiology, KU Leuven, Herestraat 49, Leuven, B-3000, Belgium

(Received 4 April 2017; revised 28 August 2017; accepted 17 October 2017; published online 27 November 2017)

Temporal effects in interaural level difference (ILD) perception are not well understood. While it is often assumed that ILD sensitivity is independent of the temporal stimulus properties, a reduction of ILD sensitivity for stimuli with a high modulation rate has been reported (known under the term binaural adaptation). Experiment 1 compared ILD thresholds and sequential-level-difference (SLD) thresholds using 300-ms bandpass-filtered pulse trains (centered at 4 kHz) with rates of 100, 400, and 800 pulses per second (pps). In contrast to the SLD thresholds, ILD thresholds were elevated at 800 pps, consistent with literature data that had previously been attributed to binaural adaptation. Experiment 2 showed better ILD sensitivity for pulse trains than for pure tones, suggesting that amplitude modulation enhances ILD sensitivity. The present ILD data and binaural adaptation data from the literature were predicted by a model combining well-established auditory periphery front-ends with an interaural comparison stage. The model also accounted for other published ILD data, including target ILD thresholds in diotic forward and backward fringes and ILD thresholds with different amounts of interaural correlation. Overall, a variety of temporal effects in ILD perception, including binaural adaptation, appear to be largely attributable to monaural peripheral auditory processing. © *2017 Acoustical Society of America*. https://doi.org/10.1121/1.5009563

[GCS] Pages: 3267–3283

I. INTRODUCTION

Spatial hearing is an important requirement for determining the location of a sound source, for following a target speaker in environments of interfering sounds, and for overall spatial awareness. The left/right dimension of spatial hearing, so-called lateralization, relies on the two binaural cues, interaural time differences (ITDs) and interaural level differences (ILDs). Most research on binaural hearing focused on the perception of ITDs, which are known to dominate the lateralization of sounds containing low frequencies (Macpherson and Middlebrooks, 2002; Wightman and Kistler, 1992). Research on ILDs, which are more important for the lateralization of higher-frequency sounds (Macpherson and Middlebrooks, 2002), is however, much more scarce. In particular, the dependency of ILD perception on temporal stimulus properties has received relatively little attention so far (Hafter *et al.*, 1983; Hafter and Buell, 1990; Stecker and Brown, 2010, 2012).

ILDs arise from the position- and frequency-dependent acoustic head shadow and vary systematically with source azimuth (Middlebrooks and Green, 1991; Xie, 2013). Besides their dominant role in lateralization of high-frequency sounds, ILDs seem to be particularly perceptually robust cues in reverberation (Devore and Delgutte, 2010; Hartmann and Constan, 2002; Rakerd and Hartmann, 2010), they contribute to spatial unmasking of speech in interfering sounds (e.g., Kidd *et al.*, 2010), and they represent the most salient sound localization cue in cochlear implant listeners (Grantham *et al.*, 2008; Laback *et al.*, 2004; Majdak *et al.*, 2011; Seeber and Fastl, 2008). Despite these important contributions of ILD little is known about how temporal envelope fluctuations, an important property of most natural sounds, affect the perception of ILDs and if known stages of auditory processing can account for temporal effects in ILD perception. These issues are the focus of the present study.

In a series of experiments, Hafter and colleagues (Hafter *et al.*, 1983; Hafter and Buell, 1990; Hafter and Dye, 1983) reported that the sensitivity to both ITDs and ILDs decreases with increasing repetition rate (thus, modulation rate) of trains of high-frequency bandpass filtered pulses when their rate exceeds about 100–200 pulses per second (pps). They concluded that with increasing rate ITD and ILD cues in the ongoing signal (after the onset) contribute less to binaural sensitivity. Later studies using similar stimuli, however, found better sensitivity to ILD in the onset and offset than in the interior pulses (Stecker and Brown, 2012) or even no evidence for rate-limited or onset/offset-dominated ILD

[a)]Parts of this work were presented at the 169th meeting of the Acoustical Society of America in Pittsburgh.

[b)]Electronic mail: Bernhard.Laback@oeaw.ac.at

[c)]Current address: School of Communication Sciences and Disorders, Western University, London, Ontario N6G 1H1, Canada.

processing (Stecker and Brown, 2010), although the latter finding may partly be due to the more restricted ranges of rates tested. Common to all the abovementioned ILD studies is that the comparisons across rates were made using stimuli with a constant number of pulses. One question left open by those studies is, therefore, how pulse rate affects ILD sensitivity for stimuli with a constant duration rather than a constant number of pulses, a comparison which is probably more relevant with respect to practical applications (e.g., stimulation strategies for cochlear implants). Theoretically, considering each pulse as independent information unit, the larger number of pulses at higher rates for constant-duration stimuli might outweigh rate effects observed for stimuli with a constant number of pulses across rates. Comparing thresholds from Hafter *et al.* (1983) for available conditions with an approximately similar duration across rates ($\sim$20–30 ms) suggests better sensitivity at intermediate rates (200 and 500 pps) than at lower and higher rates, thus a nonmonotonic rate effect. In experiment 1 of the present study we measured ILD thresholds as a function of pulse rate while keeping the stimulus duration constant. By using a stimulus duration of 300 ms, we attempted to provide conditions where temporal integration is assumed to be saturated (Buell and Hafter, 1988). A second question is if rate effects are specific to binaural hearing or if they also operate on sequential level comparison (sequential-level-difference, SLD). Several studies showed better sensitivity in ILD tasks than in SLD tasks (Hafter *et al.*, 1977; Hartmann and Constan, 2002; Stellmack *et al.*, 2004). To compare the effect of rate in these tasks, in experiment 1 we also measured SLD thresholds across pulse rates and compared them to the ILD thresholds. A third question is to what extent the presence of amplitude modulation (AM), irrespective of its particular rate, affects the perception of ILD. In a perceptual learning study Zhang and Wright (2009) reported lower ILD thresholds for sinusoidally amplitude modulated (SAM) tones than for pure tones. They noted that temporal fluctuations may be beneficial for ILD coding but did not further address the underlying mechanisms. To investigate the effect of modulation on ILD sensitivity in more depth, experiment 2 compared ILD thresholds between pulse trains and an unmodulated pure tone with a frequency matched to the spectral center of the pulse trains.

The second focus of this study is on the mechanisms underlying temporal effects in ILD perception. Hafter and colleagues (Hafter, 1997; Hafter and Buell, 1990) suggested that the reduced sensitivity to post-onset ILD as well as ITD cues at higher rates may be due to a common mechanism that reduces the perceptual weight of ongoing binaural cues particularly at high rates. They named this effect binaural adaptation. Later, they also suggested the presence of a change or novelty detector (Hafter, 1997; Hafter and Buell, 1990), which controls the weight of post-onset binaural cues at high rates depending on the novelty in the ongoing monaural auditory inputs; accordingly, only binaural cues corresponding to novel monaural input receive a high weight. Hafter (1997) proposed the site of the novelty detector to be somewhere after the auditory nerve (AN) and before binaual interaction, and suggested the cochlear nucleus as a possible candidate site. The enhanced perceptual weight of late-arriving (offset) ILD cues has been suggested to be due to some mechanism beyond binaural interaction (Le Goff *et al.*, 2012; Stecker and Brown, 2012). Alternatively, such temporal effects in ILD perception might also be attributable to known nonlinear and dynamic properties of peripheral processing up to the level of the AN, e.g., basilar membrane (BM) nonlinearity and AN adaptation.

Hartmann and Constan (2002) tested a basic model of ILD perception that is based on the integration over time of the left and right ear signals separately and subsequent comparison of integrated levels across ears to evaluate the ILD (referred to as the level-meter model). According to this model, the temporal structure of the stimulus has no impact and only the long-time averaged ILD determines ILD sensitivity. The authors found, however, slightly but significantly lower ILD thresholds for interaurally uncorrelated compared to interaurally correlated broadband or low-pass filtered noise stimuli, showing that the temporal stimulus structure actually affects ILD sensitivity. A variant of the level-meter model proposed by the same authors (called the loudness-meter model) that incorporates basic stages of peripheral auditory processing was shown to account for the results when assuming temporal integration up to 300 ms (the stimulus duration was 500 ms). Interestingly, a model version without auditory processing stages but including temporal integration also accounted for the results when the integration time was set to 240 ms.

The level meter model would predict no effect of the presence and of the rate of modulation on ILD thresholds because stimulus variability is perfectly correlated between the two ears. Nonlinear peripheral processing, as incorporated in the loudness meter model, could, however, result in predicted ILD thresholds depending on the modulation properties if it is assumed that the level difference decorrelates the internal representations of the left and right ear signals and decorrelation, in turn, degrades ILD sensitivity.

Although a number of ILD models have been proposed (Breebaart *et al.*, 2001; Brown, 2012; Brown and Tollin, 2016; Bures and Marsalek, 2013; Hartmann and Constan, 2002; Lindemann, 1986; Reed and van de Par, 2015; Stern *et al.*, 1988; Takanen *et al.*, 2014), these models mostly did not address the impact of temporal signal properties on ILD perception. The approaches by Brown (2012) and Brown and Tollin (2016) appear particularly interesting with respect to this because they used a physiologically plausible front-end model (Zilany *et al.*, 2014), which has been shown to account for temporal aspects in monaural auditory processing up to the AN. Although Brown (2012) performed no quantitative threshold predictions, model responses were qualitatively consistent with a number of basic effects in ILD perception, including the dominance of onsets in sound lateralization (referred to as precedence effect). Brown and Tollin (2016) predicted the effects of interaural decorrelation on ILD thresholds. In the present study we built on the general approach of these models to, first, quantitatively simulate the results of experiments 1 and 2 on the presence and rate of AM in ILD perception and, then, predict ILD data from three ILD studies in the literature involving different variations of temporal stimulus properties. As the model predictions will show, the temporal effects in ILD perception considered in our study appear to be largely due to known peripheral auditory processing stages.

II. EXPERIMENT 1

A. Listeners

Seven normal-hearing listeners participated in both experiments. The first author was one of the listeners. The listeners' ages ranged from 22 to 42 yr, with a mean age of 26 yr. All listeners had thresholds of 15 dB hearing level (HL) or lower at octave frequencies from 125 to 8000 Hz (ANSI, 1996) and had previous experience in psychoacoustical experiments. The listeners were paid for their service on an hourly basis.

B. Stimuli and apparatus

The stimuli were 300-ms (including ramps, see below) pulse trains, generated by bandpass-filtering trains of monophasic pulses with a duration of 21 μs, corresponding to one sampling interval at a sampling rate of 48 kHz. Pulse trains with rates of 100, 400, and 800 pps were tested.

In the ILD experiment, an ILD (in dB) was applied by attenuating the level at one ear by half of its amount and amplifying the level at the other ear by half of its amount, relative to the nominal level [see Fig. 1(a)]. Thus, the nominal level of the dichotic stimulus corresponded to the average level (in dB) across ears. In the SLD experiment, the signal levels were identical at the two ears, i.e., the stimulus was diotic, and the nominal level corresponded to the average level across intervals [see Fig. 1(b)].

The bandpass filter used to filter the monophasic pulse trains consisted of a digital eighth-order Butterworth filter centered at 4 kHz. The −3-dB/octave cutoff frequencies of the filter were 3200 and 5000 Hz. This bandwidth was chosen so that for the highest pulse rate (800 pps) the first lower sideband of the main component at 4 kHz just fell into the filter passband. After the filtering, raised-cosine onset and offset-ramps of 50-ms duration were applied. Because of the high center frequency the filtered pulse trains were assumed to provide only envelope ITD cues.

The level of the stimuli was specified in terms of the nominal level, as described in the preceding paragraph. Given that

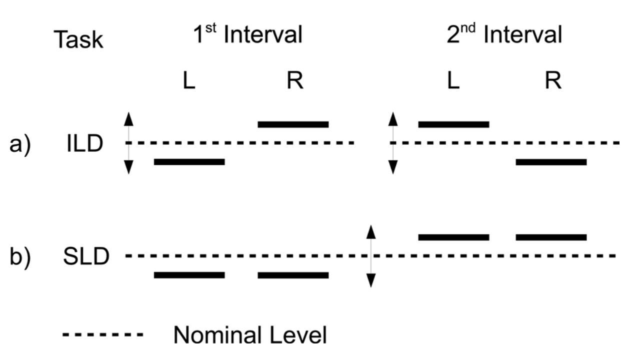

FIG. 1. Schematics of the two stimulus levels in the two intervals of a trial in the ILD and SLD tasks. The dashed horizontal line indicates the nominal stimulus level and the vertical arrows indicate roving of that level either across intervals (ILD task) or across trails (SLD task).

the root-mean-square (rms)-level depends on the pulse rate, the amplitudes of the stimuli were adjusted to maintain a constant rms-level for all rates. A pulse train at the nominal level produced an A-weighted sound pressure level (SPL) of 54 dB (re 20 μPa). Because in an ILD experiment listeners could potentially base their left/right judgments on monaural level cues (Laback *et al.*, 2004; Yost and Dye, 1988), it was attempted to preclude listeners from using such monaural cues by randomly roving the level from interval to interval. The level distribution was rectangular, centered around the nominal level, and had a width of 10 dB. According to Green (1988), for this roving width and the experimental paradigm used (see Sec. II C, Procedure) the expected threshold for decisions based on monaural level difference alone is 3.6 dB, therefore, ILD thresholds below this value cannot be based on monaural level cues. The same range of 10 dB level roving was applied in the SLD experiment. However, in the SLD experiment the roving was applied from trial to trial because interval-to-interval roving could not logically be applied because it would have disrupted the level difference to be detected. Because level roving could potentially impede the overall performance by distracting the listener, also a test condition with a fixed level was included. In the fixed-level condition the stimuli were always presented at the nominal level. One listener completed only the fixed-level condition. He was tested last, after the results for the other six listeners had shown no systematic differences between the roved-level and fixed-level conditions.

Despite the filtering of the pulse trains and the overall low stimulus level, some low-frequency nonlinear distortion products arising from normal, nonlinear peripheral auditory processing (Goldstein, 1967) might be available as discrimination cues. To preclude listeners using such cues, a binaurally uncorrelated background noise was continuously played throughout the testing. It consisted of Gaussian white noise ranging from 50 Hz to 1300 kHz with an overall A-weighted SPL of 52 dB. This corresponds to a sound pressure spectrum level of 21.1 dB (re 20 μPa in a 1-Hz band).

A personal computer system was used to control the experiments and generate the stimuli. Stimuli were output at a sampling rate of 48 kHz and a 24-bit resolution and presented via circumaural headphones (HDA200, Sennheiser, Wedemark-Wennebostel). The stimulus levels were calibrated using an artificial ear (B&K 4153, Brüel & Kjær, Copenhagen). The experiments were performed in a double-walled sound booth.

C. Procedure

The basic paradigm was a two-interval, two-alternative, forced choice (2-AFC) procedure. Each trial consisted of two 300-ms observation intervals separated by 400 ms. Stimulus intervals were marked visually using a computer monitor, starting 200 ms before the first interval and ending 200 ms after the second interval. Visual feedback on the correctness of responses was provided 200 ms after the listener's response.

In the ILD experiment, the first interval contained a dichotic stimulus with an ILD favoring one ear and the second interval contained the same stimulus but switching the signals betweeen ears. The subjects were requested to indicate whether the perceived image position moved from left to right or from right to left, by pressing an appropriate button. The direction of ILD change was random with equal *a priori* probabilities. Because the listeners are expected to base their left/right decisions on the change in ILD across intervals, each pointing to opposite directions, the nominal ILD used for threshold specification corresponds to the absolute value of the difference between the opposing ILDs presented in the two intervals.

In the SLD experiment, the stimuli were diotic and differed only in level across the two intervals. The subjects were asked to indicate which of the two intervals contained the louder stimulus. The allocation of quieter and louder stimuli to the first and second intervals was random with equal *a priori* probabilities.

The magnitude of the (interaural) level difference was varied in a staircase fashion. The staircase followed the one-up three-down rule, targeting the 79% correct point on a psychometric function. For every incorrect response the ILD/SLD was increased, and after three successive correct responses the ILD/SLD was decreased. At the beginning of an experimental run the ILD/SLD was 10 dB and the stepsize was 4 dB. The stepsize was halved after the third, fifth, seventh, and eight reversals, resulting in a final stepsize of 0.25 dB. A staircase run was finished after 12 reversals, and a threshold estimate was obtained by averaging the levels reached at the last 8 reversals. The duration of runs was, on average, about 3 min. A minimum of six threshold estimates was obtained for each condition and listener. The final threshold was defined as the average across all estimates.

Before commencing the experiment the listeners were familiarized with the experimental procedures using written instructions supplemented by oral briefing. All listeners completed at least two training runs for both the ILD and SLD tasks. In each experimental block six adaptive staircases were tested, comprising two repetitions of three pulse rates. Within one block, only one type of experiment (ILD with roved level, ILD with fixed level, or SLD) was tested. The order of blocks was randomized for each listener. After completion of six staircases for each test condition, the data were inspected for the occurrence of systematic learning effects or outliers. Only in few cases it appeared necessary to collect further staircases. Finally, all staircases collected (including the additionally collected ones) were considered as valid and were used for averaging.

The experimental blocks were spread over three to six days, depending on the availability of the listeners.

D. Results

The pattern of results was found to be similar across the seven listeners tested. Thus, we focus on the description of the average data across listeners. Figure 2 shows the across-listener averaged data as a function of the pulse rate, with error bars representing 95% confidence intervals. ILD thresholds in the fixed-level condition are denoted with blue squares, ILD thresholds in the roved-level condition with red circles, and SLD thresholds (with level roving) with green triangles. The ILD thresholds were very similar for the conditions with and without level roving, consistent with several previous studies (Bernstein, 2004; Grantham, 1984; Hartmann and Constan, 2002; Stellmack *et al.*, 2004). Furthermore, the ILD thresholds are all considerably smaller than the 3.6-dB threshold limit expected for monaural decisions based on level alone. Together, all these results strongly suggest that the listeners indeed based their responses on ILDs and not on monaural level cues.

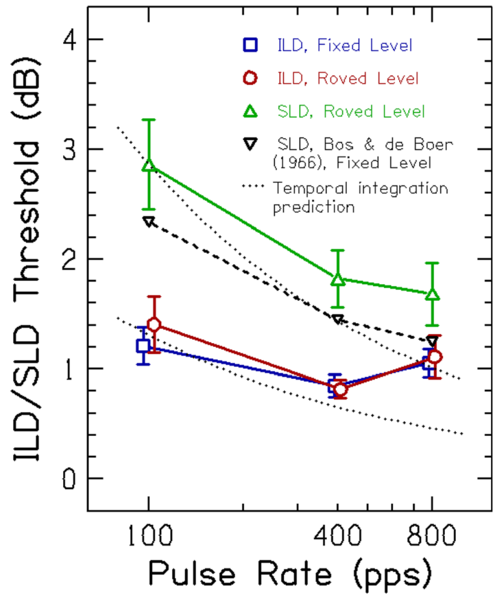

FIG. 2. (Color online) Results of experiment 1; thresholds for ILD (blue squares for fixed level, red circles for roved level) and SLD (green upward pointing triangles) as a function of pulse rate. Error bars show 95% confidence intervals. Downward pointing triangles show SLD data from Bos and de Boer (1966). The dotted lines show predictions assuming perfect temporal integration of ILD information, normalized at 100 pps (see Sec. II D for details).

The ILD thresholds decline from 100 to 400 pps and increase again from 400 to 800 pps. This suggests that two different processes are operating at rates below and above 400 pps. The SLD thresholds are all higher than the ILD thresholds at the corresponding rates. The SLD thresholds decline in the range from 100 to 400 pps, similar to the ILD thresholds. However, in contrast to the ILD thresholds the SLD thresholds saturate from 400 to 800 pps. The data were subjected to a repeated measures analysis of variance (RM ANOVA), using the factors pulse rate and experimental condition (ILD-fixed-level, ILD-roved-level, and SLD). The main effects of both factors (pulse rate: $F_{2,12} = 29.0$, $p < 0.001$; condition: $F_{2,12} = 86.6$, $p < 0.001$) and their interaction ($F_{4,24} = 5.0$, $p < 0.001$) were significant. To confirm the lack of an effect of level roving for ILD, a separate RM ANOVA was performed on the two ILD conditions, revealing a significant effect of the pulse rate ($F_{2,12} = 19.0$, $p < 0.001$), no main effect of the condition ($F_{1,6} = 0.001$, $p < 0.97$), and no interaction between them ($F_{2,12} = 0.7$, $p < 0.52$). This indicates that the interaction in the first ANOVA contrasted the ILD and SLD conditions. Given the lack of an effect of level roving, the data for the fixed- and roved-level conditions were pooled in the subsequent analyses.

Next, given the nonmonotonic effect of the pulse rate, separate RM ANOVAs were performed on the data for pulse rates ≤400 pps and ≥400 pps. For the lower rate range (100 vs 400 pps), both the main effects (pulse rate: $F_{1,6} = 54.1$, $p < 0.0001$; condition: $F_{1,6} = 153.8$, $p < 0.001$) and their interaction ($F_{1,6} = 5.8$, $p < 0.019$) were significant. Thus, while both the ILD and SLD thresholds improve from 100 to 400 pps, the improvement appears to be larger for the SLD thresholds. However, this should be considered with caution, given that a logarithmic scaling of the y axis would remove that difference and it is not clear what the appropriate scaling is. The significance of the improvement for the ILD thresholds was confirmed by a paired t-test on the ILD data only [$t(19) = 36.5$, $p < 0.001$].

For the higher rate range (400 vs 800 pps) the factor pulse rate had no significant effect ($F_{1,6} = 0.5$, $p < 0.49$), while both the factor condition ($F_{1,6} = 103.8$, $p < 0.001$) and the interaction pulse rate × condition ($F_{1,6} = 5.5$, $p < 0.02$) were significant. Subsequent t-tests showed a significant *improvement* with increasing pulse rate for the ILD thresholds [$t(14) = 14.4$, $p < 0.001$] and no effect of the pulse rate for the SLD thresholds [$t(14) = 0.5$, $p = 0.48$].

E. Discussion

The main hypothesis of this study was that the sensitivity to ILD decreases with increasing pulse rate above 100 pps, due to an adaptation mechanism operating specifically on binaural cues, namely, binaural adaptation (Hafter *et al.*, 1983; Hafter and Dye, 1983). This hypothesis implies that such a binaural rate limitation does not operate on monaural (or diotic) sequential level discrimination. From mid to high pulse rates (400–800 pps) our results show decreasing ILD sensitivity and no such effect for SLD sensitivity, consistent with the binaural adaptation hypothesis. However, from low to mid pulse rates (100–400 pps) our results show the opposite effect, i.e., an *improvement* in both ILD and SLD sensitivity. Therefore, below about 400 pps some other rate-dependent mechanism obviously affects the sensitivity to both interaural and sequential level differences.

Hafter *et al.* (1983) measured ILD thresholds for trains of pulses at rates of 100, 200, 500, and 1000 pps. For the maximum number of pulses they tested (32), they found constant sensitivity at 100 and 200 pps and decreasing sensitivity for further increases of the pulse rate. That decrease in sensitivity is qualitatively consistent with the decrease observed from 400 to 800 pps in our data. However, their data do not show an improvement in sensitivity from very low (100 pps) to mid rates (500 pps) as we observed. With respect to that second effect, an important difference in the stimuli used in the two studies should be considered. While Hafter *et al.* (1983) used a constant number of pulses, i.e., information units, across rates, our stimuli had a constant duration across rates and, hence, an increasing number of information units with increasing rate. As mentioned in the Introduction, comparing the available data of Hafter *et al.* (1983) for a constant duration across rates (∼20–30 ms) actually shows better sensitivity at intermediate rates of 200 and 500 pps than at lower (100 pps) and higher (1000 pps) rates, thus, consistent with the nonmonotonic effect of rate on ILD thresholds found in our data despite the largely differing durations across studies. With respect to our data, we hypothesized that the reason for the improvement from 100 to 400 pps is due to temporal integration across a larger number of information units (i.e., pulses). To evaluate this idea, the dotted curve in Fig. 2 shows the prediction of thresholds as expected when assuming perfect integration of information provided by each pulse in the pulse trains, assuming statistical independence of internal noise (Hafter and Dye, 1983; Hougast and Plomp, 1968). The model was set to coincide at 100 pps with the average ILD threshold across the fixed- and roved-level conditions. As can be seen, the relative performance at 400 pps is reasonably well predicted by the integration model (the slight discrepancy is probably due to binaural adaptation just starting around 400 pps), while the relative performance at 800 pps is clearly worse than predicted. Therefore, it appears that the improvement from 100 to 400 pps for our constant-duration stimuli is due to the integration across an increasing number of information units.

Another study by Stecker and Brown (2010) used very similar stimuli as Hafter *et al.* (1983) at pulse rates of 100, 200, and 500 pps, reporting no systematic effect of pulse rate on ILD sensitivity. The lack of an increase in sensitivity at low rates is likely due to the same reason as described above for Hafter *et al.* (1983), given that Stecker and Brown (2010) also used a constant number of pulses across rates. The lack of a decline of sensitivity at high rates could be due to their choice of testing only up to a maximum rate of 500 pps.

While we are not aware of any study testing the effect of varying the rate of modulation on SLD sensitivity, a study by Bos and de Boer (1966) on SLD using bands of noise with various bandwidths and center frequencies can be compared to our study. In this comparison we assume that the average rate of random envelope fluctuation for Gaussian white noise corresponds to ∼64% of its bandwidth (Rice, 1945). The black downward-pointing triangles in Fig. 2 show SLD thresholds extracted from the data of Bos and de Boer (1966) using that envelope rule, averaging across center frequencies and interpolating across noise bandwidths actually tested. Despite a 0.5-dB offset, the rate dependency is remarkably similar between the two studies, and consistent with our finding of improving sensitivity up to intermediate rates. The dotted line in Fig. 2 shows the prediction of SLD thresholds assuming perfect integration of information across pulses. The relative SLD performance is slightly worse than predicted at 400 pps and clearly worse than predicted at 800 pps. Thus, similar to the ILD data the improvement is slightly less than optimal up to 400 pps and considerably less than optimal up to 800 pps. Taken together, it appears plausible that the mechanism causing improving sensitivity from low to mid rates is the same for the ILD and SLD tasks, namely, temporal integration.

Our results show clearly better sensitivity to ILDs than to SLDs. The difference between SLD and ILD thresholds, averaged across pulse rates and across fixed and roved-level conditions, amounts to 1.01 dB. This suggests that the auditory system is more sensitive in detecting changes in level across ears than successive changes in level at each ear.

Hafter *et al.* (1977) compared ILD and monaural SLD sensitivity using high-frequency-filtered pulses in a two-interval same/different task. Using our threshold specification in terms of nominal ILD, they found a very similar threshold difference of 1.0 dB. A study by Stellmack *et al.* (2004) compared ILD sensitivity in a single-interval task with SLD sensitivity in a two-interval task, reasoning that these two tasks are comparable in terms of the complexity of the information presented to the listener. They found lower thresholds in the ILD task for broadband noise stimuli. However, it is possible that the comparison was confounded by memory constraints involved in sequential level comparison in the SLD task as compared to simultaneous level comparison in the ILD task (Hartmann and Constan, 2002). This might have led to relatively higher thresholds in the SLD task. In addition, listeners might have used monaural absolute level cues in the ILD task. Hartmann and Constan (2002) reported ILD and SLD thresholds using broadband and low-pass-filtered noise stimuli. On average across their three experiments and considering only interaurally correlated stimuli, they reported a difference between SLD and ILD thresholds of 0.35 dB. In fact, the comparison appears to be complicated by different definitions of ILD used in the two studies. While we defined the ILD as twice the absolute value of the ILDs presented in the two intervals of a trial (i.e., the change in ILD across intervals), Hartmann and Constan (2002) defined the ILD as present in each of the two intervals. Multiplying their ILD thresholds by two to make their ILD definition compatible with ours leads to very similar ILD thresholds (our study: 1.06 dB; their study: 1.26 dB). However, with this definition their listeners' ILD thresholds become higher (worse) than their SLD thresholds (0.98 dB), inconsistent with our study and the other studies described. A possible explanation could be a difference in the design of the SLD experiment. While Hartmann and Constan (2002) used a fixed signal level, we used random level roving across trials in an attempt to produce conditions more comparable to the ILD condition with level roving. Although we observed no effect of level roving in the ILD experiment, level roving may have raised the thresholds in the SLD experiment. Pienkowski and Hagerman (2009) showed that adding a moderate amount of level roving increases SLD thresholds by about 0.38 dB on average across listeners. Applying this correction to the SLD thresholds reported by Hartmann and Constan (2002), this would amount to SLD thresholds of 1.36 dB, still only minimally higher than the corresponding ILD thresholds. It is, therefore, not clear why Hartmann and Constan did not find a larger difference between SLD and ILD thresholds.

Finally, the question arises if the binaural system is indeed more accurate than the monaural system in extracting level information. While both the ILD and the SLD tasks involve memory constraints because of the comparison of stimuli across intervals within a trial, it is possible that the short-term memory for spatial position (ILD task) is more accurate than the short-term memory for loudness (SLD task; Hartmann and Constan, 2002). This would suggest that neural-processing demands require higher accuracy for tracking changes in sound source position based on ILD than for tracking changes in level. In any case, our data reveal that the binaural system outperforms the monaural system in evaluating level information while they do not reveal whether the better sensitivity in the binaural task is due to a higher accuracy of the binaural (instantaneous) level extraction process *per se* or to a higher accuracy in tracking changes in spatial position over time.

In summary, where comparison conditions are available the results of the present study are, for the most part, compatible with the existing literature on the dependency of ILD and SLD sensitivity on the envelope rate and on the overall difference between ILD and SLD sensitivity. The increase in sensitivity from low to mid rates can be understood in terms of an increasing number of information units across which the auditory system integrates level information. The saturation of the improvement at mid rates, as observed for the SLD data, is consistent with the idea that this mechanism is limited by the limited temporal resolution of the auditory system. The *decrease* of sensitivity from mid to high rates, as observed for the ILD task only, suggests the involvement of an additional mechanism operating on binaural information only. The nature and origin of this mechanism are studied in more depth in a follow-up experiment (experiment 2) and by means of a modeling approach (Sec. IV).

III. EXPERIMENT 2

A. Rationale

The results of experiment 1 showed that ILD sensitivity for pulse trains worsens when the pulse rate exceeds 400 pps. This result seems to suggest that ILD sensitivity worsens when the duration of envelope dips becomes very short. Following this idea an unmodulated pure tone, representing a pulse train with an infinitely high rate and, so, no envelope dips, should elicit lower ILD sensitivity than a pulse train containing envelope modulation. In experiment 2 we test the hypothesis that pure tones elicit lower ILD sensitivity than pulse trains because of the lack of envelope modulation.

B. Conditions

The stimuli were an unmodulated 4-kHz pure tone and a filtered 100-pps pulse train spectrally centered at the same frequency. Both stimuli were shaped with 50-ms raised-cosine onset and offset-ramps of 50-ms duration. The pulse train served as control condition for comparison of the results with the data from experiment 1 and had the same level, i.e., 54 dB SPL. The pure tone was presented in two different level conditions: in the equal-rms-level condition, the *rms*-level was matched to that of the pulse train. In the equal-peak level condition, the *peak* level was matched to that of the pulse train, resulting in a SPL of 67 dB. Level roving was not included given that experiment 1 showed no systematic effect of level roving. The three conditions (pulse train, equal-rms pure tone, equal-peak pure tone) were tested in block-wise randomized order across listeners. All other aspects of the experiment, including the seven listeners, were the same as in experiment 1.

C. Results

Figure 3 compares the threshold ILDs averaged across the seven listeners, including the results for the pulse trains in the condition without level roving from experiment 1 for comparison. The thresholds for the 100-pps conditions were very similar in experiments 1 and 2 (1.2 and 1.3 dB, respectively), suggesting that a direct comparison of results across experiments is justified. The ILD thresholds for the equal-rms pure tone was almost twice as high (2.4 dB) as those for the 100-pps pulse train, while the thresholds for the equal-peak pure tone were only slightly higher (1.7 dB).

A RM ANOVA was performed on the data for the three conditions, supplemented with the three pulse-train conditions (100, 400, and 800 pps) from experiment 1. The effect of the factor condition was found to be significant (pulse rate: $F_{5,42} = 18.1$, $p < 0.001$). A Tukey's *post hoc* test was performed on the six levels of the factor condition. The 100-pps conditions from experiments 1 and 2 were not significantly different ($p = 0.999$), supporting the comparability of results between the two experiments. The equal-rms pure tone differed significantly from all other conditions ($p \leq 0.001$), while the equal-peak pure tone differed significantly from the 400 and 800-pps pulse trains from experiment 1 ($p = 0.001$ and 0.014, respectively) but not from the 100-pps pulse train tested in the two experiments ($p = 0.25$).

D. Discussion

The observation of significantly higher ILD thresholds for a pure tone compared to pulse trains with various rates supports the hypothesis that the lack of envelope modulation is disadvantageous for ILD sensitivity. On a more detailed level, this effect was more pronounced when the pure tone had the same rms-level rather than peak-level relative to the pulse train, indirectly suggesting that the peak portions of the pulse train's envelope are more important for ILD sensitivity. The level effect in ILD sensitivity observed by comparing ILD thresholds between the lower-level and the higher-level pure tones (2.4 and 1.7 dB, respectively) is consistent with preceding studies, e.g., Dietz *et al.* (2013).

The beneficial effect of AM for ILD sensitivity reported here is overall consistent with data by Zhang and Wright (2009) who reported lower ILD thresholds for SAM tones than for a pure tone with the same carrier frequency of 4 kHz and the same SPL.

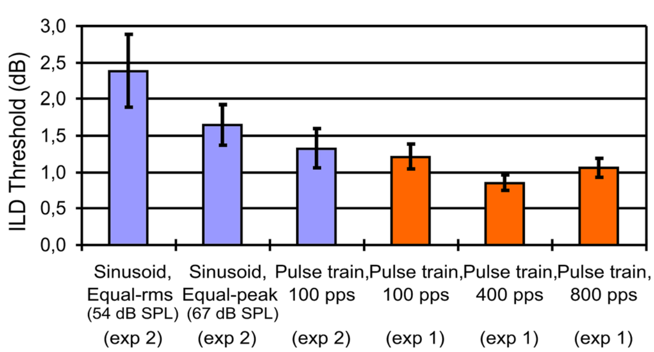

FIG. 3. (Color online) Results of experiment 2; ILD thresholds for a 100-pps pulse train and a 4-kHz pure tone at equal SPL level (54 dB SPL) and equal peak level (67 dB SPL) relative to the pulse train. No level roving was used. Error bars show 95% confidence intervals. The three right-most bars show corresponding results from experiment 1.

There is the possibility that the higher ILD sensitivity for pulse trains relative to unmodulated pure tones is due to their wider spectral bandwidth. The auditory system could take advantage of integrating information across frequency bands (or a larger number of activated neurons), as shown, for example, by improved thresholds for ITDs when the stimulus bandwidth is increased (e.g., Buell and Hafter, 1991). Figure 4 shows, for the different stimuli from experiments 1 and 2, the corresponding AN response areas (RAs), i.e., the mean AN firing rates as a function of center frequency (CF), as predicted by an AN model described in Sec. IV A. While the RAs indeed tend to be broader for the pulse trains than for the pure tones, there are a few arguments that do not support an explanation in terms of spectral integration alone. First, the RA width tends to be at least as broad for the 800-pps pulse train as for the 400-pps pulse train, although the corresponding ILD threshold was found to be significantly higher for the 800-pps pulse train. Second, the RA width is only marginally narrower for the equal-peak (67-dB SPL) pure tone than for the 400-pps pulse train, although the ILD threshold was found to be significantly higher. Third, as will be shown in Sec. IV B 1, a model of ILD sensitivity based on the AN responses largely predicted the pattern of results, although the prediction was based on only one fiber centered on the stimulus (4 kHz) and, thus, spectral integration was not accounted for. Taken together, while the different effective bandwidths of the stimuli might have contributed to the measured sensitivities, bandwidth alone does not seem to be responsible for the pattern of results obtained. This is in line with Dietz *et al.* (2013), finding no difference in ILD thresholds for transposed tones compared to SAM tones across a wide range of levels despite the wider bandwidth of transposed tones.

Combination of the results from experiments 1 and 2 suggests that the higher sensitivity for pulse trains relative to pure tones found in case of ILD may not occur with respect to SLD. While we did not include this comparison in our experiments, Buus (1990) measured SLD thresholds for pure tones and noises of various bandwidths, which appear to be relevant. The thresholds for noise bandwidths up to 683 Hz, resulting in mean envelope fluctuation rates in the order of those of the modulation rates of the pulse trains of the present study, did not significantly differ from the thresholds for pure tones. These results are consistent with our conclusion that the beneficial effect of modulation is specific to ILD perception.

IV. MODELING

The results of the present and other studies involving temporal effects in ILD perception were predicted by combining a well-established front-end model of the auditory periphery with an interaural neural discharge-rate comparison process. The overall model approach resembles a model put forward in Brown and Tollin (2016). The models differ, however, in detailed aspects, including the treatment of

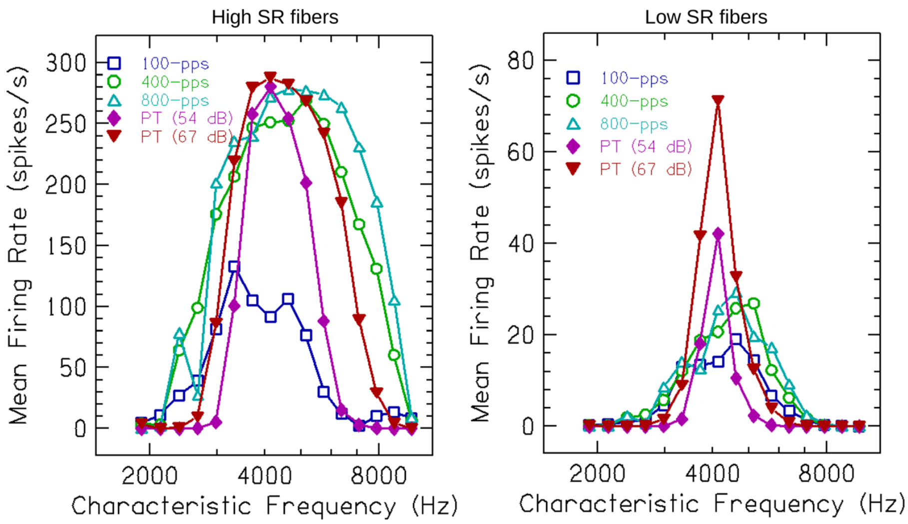

FIG. 4. (Color online) Response areas (RAs; mean firing rates across CFs predicted by the AN model by Zilany *et al.*, 2014) for stimuli used in experiments 1 and 2. The pulse trains with various rates are depicted with empty symbols and the pure tones (PT) with equal-SPL (54 dB SPL) and equal-peak level (67 dB SPL) are depicted with filled symbols.

negative values in calculating the interaural discharge rate difference, the inclusion of a discharge-rate based weighting of ILD information, or the decision statistics. In the following, we shortly summarize the front-end model, describe the interaural comparison stage in detail, and provide model predictions for the present ILD data and ILD data from the literature.

A. Model description

1. Front-end AN model

For generating AN responses to the experimental stimuli at the left and right ears, we used the latest "humanized" version (5.2) of a family of phenomenological models of the transformation of acoustic stimuli into AN discharges as described in Zilany *et al.* (2014) and Zilany *et al.* (2009) and implemented in the auditory model toolbox (Søndergaard and Majdak, 2013). Briefly, the model includes middle-ear transduction, time-varying nonlinear filters that represent physiological processes in the cochlea, a power-law synapse model, and a spike generator. This model has been shown to accurately predict a large body of AN data, including amplitude-modulation transfer functions, long-term adaptation behavior, forward masking, and adaptation to increments and decrements in the amplitude of an ongoing stimulus or tone-in-noise detection (e.g., Zilany *et al.*, 2009; Mao and Carney, 2015). Our predictions were based on model-generated probabilistic AN peri-stimulus time histograms (PSTHs, providing the instantaneous discharge rate across time while taking into account neural refractoriness) using a model sampling rate of 100 kHz. The PSTHs thus represent the instantaneous AN firing rates across time as obtained when averaging across an essentially infinite number of neurons (Zilany *et al.*, 2014). The duration of the response window was twice the signal duration (Zilany *et al.*, 2009) plus a constant of 60 ms, which was added to ensure that even for the very short experimental stimuli the responses to the target fall into the response window, given the latency in neural response. Responses of neurons at 11 different characteristic frequencies evenly spaced in the equivalent rectangular bandwidth scale (Glasberg and Moore, 1990) within the bandwidth of the particular stimuli under consideration (approximately within one octave band) were simulated and responses were averaged across those neurons. For predicting the data collected in the current study, however, we present simulations using fibers with a CF centered at the stimulus only. We found that including fibers with CFs remote from the stimulus' spectral center largely degraded the prediction performance. Regarding fiber types, predictions were made separately for the following types and their combinations: low and mid spontaneous discharge rate (SR); mid and high SR; low, mid, and high SR. In case of multiple fiber types, the responses from different fiber types were combined by linear averaging. We also tried weighted averaging according to the prevalence of different fiber types in cats (Liberman, 1978), i.e., low SR: 16%; medium SR: 23%; high SR: 61%. However, we finally skipped weighted averaging because it typically resulted in poorer predictions. In the AN response simulation, approximate implementation of power-law functions and a fixed fractional noise were used (see Zilany *et al.*, 2009). Very similar results were obtained when using actual power-law functions and variable fractional noise.

2. Interaural comparison stage

The left panel of Fig. 5 shows the characteristic sigmoidal shape of the rate-ILD function of a single lateral superior olive (LSO) cell when increasing the ipsilateral input level from values smaller than the contralateral level to values greater than the contralateral level (e.g., Tollin, 2003; Tollin *et al.*, 2008). Our model followed the idea of so-called population-code models of localization that perceptual ILD sensitivity is based on the subtraction of the outputs of left and right LSO units, presumed to take place at some higher auditory processing stage (Park *et al.*, 2004; Stecker *et al.*, 2005). There are anatomical mechanisms to accomplish this operation and the importance of this operation for the robustness against level roving has been shown (see, e.g., Tsai *et al.*, 2010). The right panel of Fig. 5 shows the difference between rate-ILD functions of left- and right-ear LSO units. This hemispheric difference function is symmetric around zero ILD and, within the range of ILDs relevant for the current study (±10 dB), approximately linear. Based on this

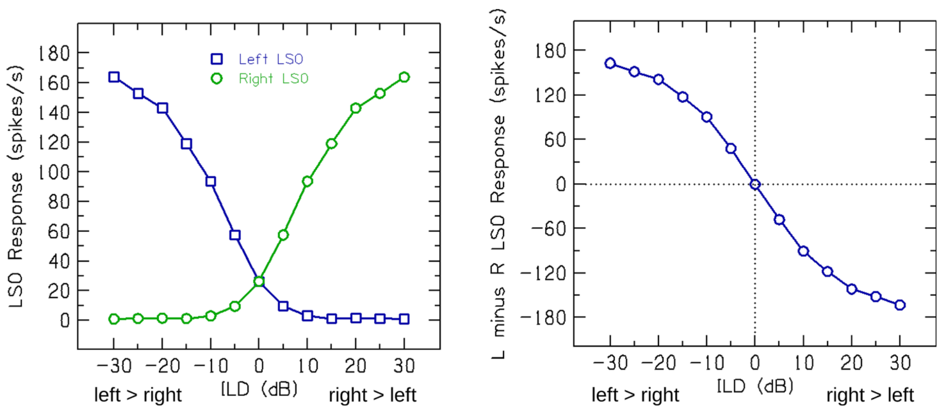

FIG. 5. (Color online) (Left) Output firing rate of LSO unit in cats in response to a range of ILDs between ipsi- and contralateral input. The squares indicate the left LSO (adapted from Tollin *et al.*, 2008) and the circles indicate the mirrored response for a hypothetical right LSO. (Right) Difference in output rate between left and right LSO units. Note the symmetry around zero ILD and approximate linearity within ±10 dB ILD.

quasi-linearity and the finding that neural sensitivity of single LSO cells to ILD is well described by a nonparametric d'-like standard separation metric (Sakitt, 1973; Tollin *et al.*, 2008), we defined the following basic ILD metric:

$$D = \frac{f(L) - f(R)}{\sqrt{\sigma(L)\sigma(R)}}, \tag{1}$$

where L is the PSTH at the left ear, R is the PSTH at the right ear, f is the mean discharge rate, and σ is the standard deviation of the discharge rate. Thus, D evaluates the difference in mean discharge rates between left and right ear AN inputs and the variability of these rates over repeated stimulus presentations. D is positive for ILDs favoring the left ear and negative for ILDs favoring the right ear.

There are indications from both physiology (Irvine *et al.*, 2001; Joris and Yin, 1995; Park *et al.*, 1996; Sanes, 1990; Tollin, 2003; Brown and Tollin, 2016) and psychophysics (Akeroyd and Bernstein, 2001; Le Goff *et al.*, 2012) that the ILD extraction mechanism does not simply integrate level information across a single long-time window covering the stimulus duration and then compare across ears, but instead reads out ILD in short windows. Therefore, in our model, referred to as the multiple short windows (MSWs) model, the left- and right-ear discharge rates from the PSTHs are determined within short rectangular windows k (with 50% overlap), resulting in L_k and R_k. The interaural differences between L_k and R_k are then averaged across all m windows, applying a weight (W_k) for each window m based on the mean discharge rate of window m across ears. Linear weighting up to a maximum discharge rate *fmax* is applied, resulting in weights of one for windows with discharge rates exceeding *fmax*. Then, the weighted standard deviations across L_k and R_k, are calculated [denoted here simply as $\sigma_w(L)$ and $\sigma_w(R)$] to finally determine D_{MSW}, i.e.,

$$D_{\mathrm{MSW}} = \frac{\left(\sum_{k=1}^{m}(L_k - R_k)W_k\right) \Big/ \left(\sum_{k=1}^{m} W_k\right)}{\sqrt{\sigma_w(L)\sigma_w(R)}}. \tag{2}$$

The duration of windows k was chosen to be 2 ms, within the range of durations proposed in the physiological literature (Irvine *et al.*, 2001; Joris and Yin, 1995; Park *et al.*, 1996; Sanes, 1990; Brown and Tollin, 2016). Weighting of the individual windows was applied to avoid that windows with discharge rates falling below a certain critical level, e.g., before stimulus onset, after stimulus offset, or during silent signal portions, would not properly encode level information and thus add noise to the estimation of the stimulus' ILD. The parameter *fmax* was chosen as a fixed percentage of the dynamic range of discharge rates for the particular fiber types used (taken from Baumgartner *et al.*, 2016). For the experiments predicted here, *fmax* was set to correspond to 32% of the neural dynamic range. Applying weighting was found to improve the predictions compared to without weighting. While a full parametric optimization of *fmax* would by far have exceeded the practically available computational resources available (particularly in combination with the other parameters), heuristic testing revealed good predictions with the chosen value.

We run all simulations also using a model variant without short-time windowing, thus integrating across the entire stimulus duration [as denoted in Eq. (1)], similar to the loudness meter model by Hartmann and Constan (2002). In that model variant, referred to as long-time-integration (LTI) model, the discharge-rate based weighting function was applied to each bin of the PSTHs. We focus here on reporting results of the physiologically more plausible MSW model.

For all simulations, the stimulus ILD always favored the left ear, i.e., L was higher than R. In order to predict ILD thresholds, first the ILD was systematically varied in steps of 0.2 dB by increasing the level of R and decreasing the level of L, each by half of the ILD magnitude (as done in the experiments), to obtain a function relating D_{MSW} (the "internal" ILD) to the stimulus ILD (applying linear interpolation). The ILD threshold was then determined at the criterion internal ILD for which predicted ILD thresholds best accounted for the experimental ILD thresholds using a variance-based metric.[1]

Extensive model simulations with many combinations of fiber types and values of the parameter *fmax* for all the data sets reported below revealed predictions tending to be best when including only high SR fibers for lower-level stimuli, low SR fibers for higher-level stimuli, and all three fiber types for mid-level stimuli.

B. Model predictions

1. Present study

First, the model was evaluated on the experimental results of the present study, using ILD thresholds for pulse trains without level roving from experiment 1 and for the constant-SPL pure tone from experiment 2. Figure 6 shows the model responses to the 100-, 400-, and 800-pps pulse trains, as well as for the pure tone (from top to bottom rows) for a fixed ILD of 1.4 dB. The left column shows PSTHs for the left and right ears (lighter and darker lines, respectively; note that the difference between left and right PSTHs is very small and can hardly be seen in this plot). Therefore, the middle column shows zoomed views of left- and right-ear AN PSTHs from left column at a temporal position indicated with the arrow at the top of Fig. 6. The right column shows the difference between left and right ear PSTHs within short-time windows (as used in the MSW model). The monaural neural responses (left column) show constant response magnitude for the 100-pps condition (apart from onset/offset fading due to the stimulus ramps), whereas all other conditions show onset dominance, which appears to be related to AN adaptation. In the zoomed view around the first response peak (middle column) it can be seen that there are instances where the response at the ear with the lower-level (dashed red line) has as larger response than the higher-level ear (solid blue line). The interaural response differences within 2-ms windows (right column) show onset dominance, with much larger short-term interaural response difference at the onset than across the ongoing stimulus. The 100-pps condition, however, differs from the other conditions by yielding weaker onset dominance. Remarkably, the ongoing responses overall carry mostly only weak interaural differences (see insets in right column). While these ongoing responses are, on average, even negative for the 100-pps condition, they are more positive for the 400-pps condition and close to zero for the 800-pps and pure tone conditions.

Figure 7 compares the experimental ILD thresholds (empty squares) with the ILD thresholds predicted by the MSW model (filled triangles). The model correctly predicts the presence of a nonmonotonic pattern of thresholds, although neither of the two models could predict a threshold for the 100-pps condition (even for very large external ILDs). Considering the predictable thresholds the model accounts for 86.3% of the variance in the experimental thresholds. In case of the 100-pps condition the model appears to lack some information the listeners had access to.

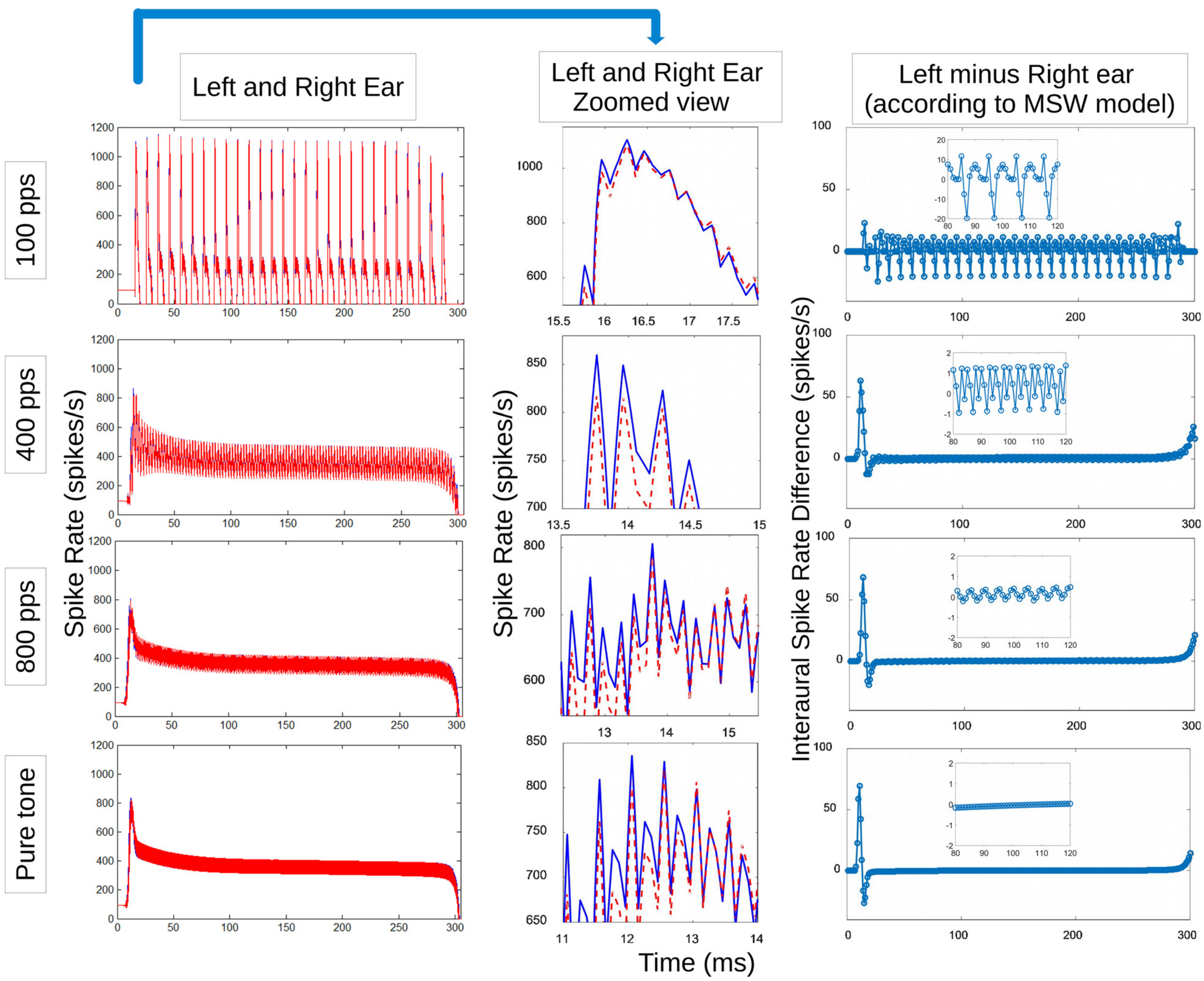

FIG. 6. (Color online) (Left column) Left- and right-ear AN PSTHs for the pulse train at three rates and equal-SPL pure tone. Note that the difference between left and right PSTHs is very small and can hardly be seen in this plot. (Middle column) Zoomed view of left- and right-ear AN PSTHs from left column about at temporal position indicated with arrow at top left of figure. The solid blue line shows the left ear response, the dashed red line shows the right ear response. (Right column) Difference between left and right ear PSTHs after windowing in 2-ms windows (according to MSW model). These differences thus represent short-time ILD representations in the model. The insets show zoomed views of a segment from the ongoing response. See Sec. IV B 1 for more details.

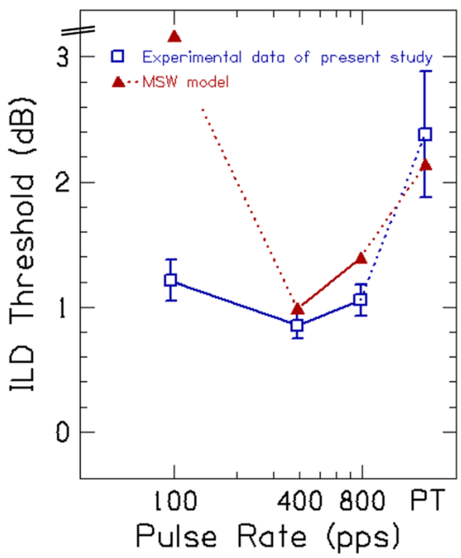

FIG. 7. (Color online) ILD thresholds for pulse trains at 100, 400, and 800 pps (from experiment 1) and equal-SPL pure tone (PT, from experiment 2), indicated with empty squares, are compared with predictions of the MSW model (filled triangles connected with dashed line). The 100-pps thresholds are not predictable within the evaluated range of ILDs (up to 6 dB) and are annotated at the upper limit of the figure.

Accordingly, Fig. 6 shows a considerable proportion of instantaneous "internal ILD" pointing to the "wrong" side for that stimulus (see corresponding inset in upper right panel), which leads to a slightly negative D_{MSW} and, thus, prevents the estimation of a valid threshold. Closer inspection of the short-time window outputs revealed that the mean ongoing discharge rate (thus, the numerator of the D_{MSW} metric) determined the pattern of predicted thresholds across stimulus conditions, whereas the standard deviation (the denominator of the D_{MSW} metric) remained approximately constant. For the 400-pps condition, the dominance of positive peaks caused highest predicted sensitivity and, in contrast, for the 100-pps condition the dominance of negative peaks caused lowest predicted sensitivity. These negative peaks appear to be related to the level-dependent neural response latency, which is most pronounced at 100-pps. Note that while the time constant of the windows in the MSW model can interfere with the time constant of neural response latency, a model version with long time constants, corresponding to the entire signal duration, predicted the same pattern of results. In Sec. IV C we discuss the issue of response latency further and propose an explanation how human listeners might extract ILD cues at the 100-pps rate. Finally, the reduction of ILD sensitivity from the 400-pps to the 800-pps and pure-tone conditions appears to be related to the reduction of post-onset interaural response difference. This corresponds well to the finding of reduced weighting of post-onset ILD for pulse rates exceeding 400 pps (Brown and Stecker, 2010), which has been associated with binaural adaptation (Hafter *et al.*, 1983).

2. *Hafter et al. (1983)*

Given the ability of the model to predict the nonmonotonic effects of the modulation rate, we next tested to what extent it accounts for the effects of systematic variation of the signal duration (in terms of the number of stimulation pulses) and pulse rate on ILD thresholds as tested in Hafter *et al.* (1983). The pulse trains in that study were bandpass filtered at 4000 Hz. Figure 8 shows the experimental data for different pulse rates (empty symbols connected with solid lines; see legend) as a function of the number of pulses in each train. As expected from temporal integration, the ILD thresholds decreased with increasing number of pulses. But more importantly, the improvement due to temporal integration became weaker for rates >200 pps, as reflected by the decreasing slopes of the curves with increasing rate (representing the binaural adaptation effect). This can be seen by the decreasing slope of the functions with increasing pulse rate. The MSW model (filled symbols connected with dotted lines in Fig. 8) is able to explain 82.3% of the variance. The model captures the main effects of the data with the most marked exception that the predicted thresholds for the 100–pps condition are much higher than the actual data, particularly for larger numbers of pulses. This is consistent with the predictions for the data of the present study described in Sec. IV B 1 (and discussed further in Sec. IV C). The reason for being able to predict 100-pps thresholds here, in contrast to the data of the present study, turned out to be that the stimuli of Hafter *et al.* contained no ramps, resulting in an enhanced onset ILD cue. In summary, apart from the 100-pps condition, both model variants are able to well predict rate-dependent temporal integration of ILD information.

3. *Le Goff et al. (2012)*

Le Goff *et al.* (2012) studied the effect of temporally surrounding a target stimulus carrying ILD by diotic stimuli. Target durations were 5, 10, 20, 50, or 100 ms, each combined with fringe durations of 5 or 100 ms. The results were considered as a measure of the shape of the temporal window for ILD perception. Both the target and the fringes were interaurally correlated wide-band noises bandpass filtered from 100 to 2900 Hz. For the simulations, four independent

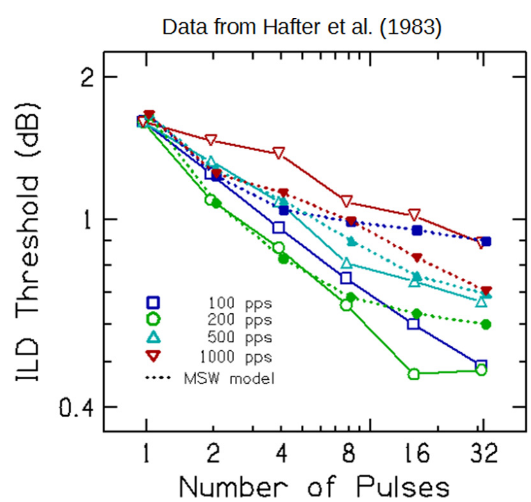

FIG. 8. (Color online) Empty symbols connected with solid lines: ILD thresholds as a function of the number of pulses in a train for various pulse rates (see legend), adapted from Hafter *et al.* (1983). Filled symbols connected with dashed lines: Predictions of the MSW model. Data and predictions are slightly shifted horizontally for the sake of clarity.

noise tokens were generated for each condition and each ILD. The responses of AN neurons at 11 different CFs from 381 to 762 Hz were simulated. Figure 9 shows the data and MSW model predictions for 5-ms fringes. The MSW model accounts for 85.7% of the variance in the data, capturing all main aspects of the data, i.e., the overall decrease in thresholds with increasing target duration (temporal integration), although to a smaller extent than the experimental data, the elevation of thresholds when adding noise fringes, being much stronger for short than for long target durations, and the stronger threshold elevation for surrounding fringes than for either forward or backward fringes. The only marked deviation is the smaller elevation of predicted thresholds for the forward fringes at short target durations compared to the experimental data. Figure 10 shows the data and model predictions for 100-ms fringes. For those data, both models provided poor predictions when the conditions with and without noise fringes were all predicted with the same criterion (D_{MSW}). When optimizing D_{MSW} separately for the noise fringe conditions the model accounted for 73.0% of the variance. This percentage excludes three data points for very short targets where thresholds could not be determined within the evaluated ILD range (up to 8 dB), and are indicated with upward pointing arrows. The measured ILD thresholds for those conditions amounted to about 5–7 dB. The overall pattern of thresholds, particularly the largely differing patterns across different conditions, is predicted by the model, most notably the stronger threshold elevation for the backward and surrounding fringes than for the forward fringes, particularly at the shorter target durations. The stronger effect of backward fringes than of forward fringes has been attributed to the increasing importance of binaural information toward the stimulus offset ("upweighting" effect; see Stecker and Brown, 2012). The ability of the model to predict this somewhat unexpected pattern of thresholds is remarkable. Le Goff *et al.* (2012) could predict these results only when combining a temporal weighting model

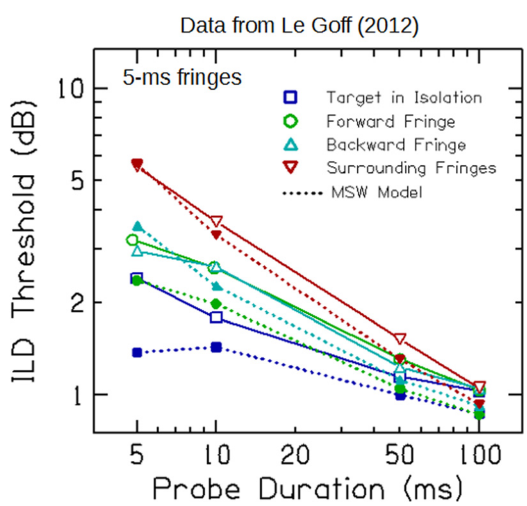

FIG. 9. (Color online) Empty symbols connected with solid lines: ILD thresholds as a function of the target duration in isolation and for various configurations of 5-ms diotic fringes (see legend), adapted from Le Goff *et al.* (2012). Filled symbols connected with dashed lines: Predictions of the MSW model.

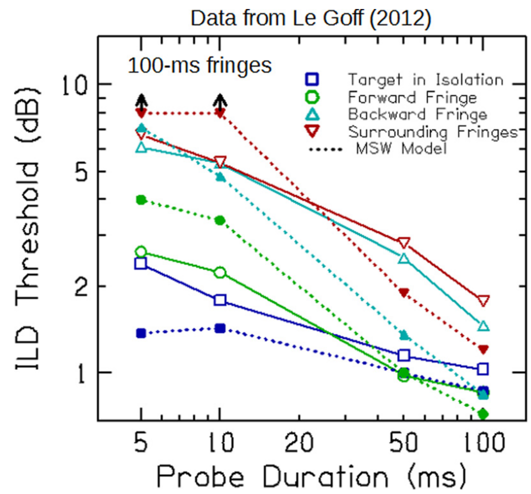

FIG. 10. (Color online) Empty symbols connected with solid lines: ILD thresholds as a function of the target duration in isolation and for various configurations of 100-ms diotic fringes (see legend), adapted from Le Goff *et al.* (2012). Filled symbols connected with dashed lines: Predictions of the MSW model.

(Akeroyd and Bernstein, 2001; Bernstein *et al.*, 2001) with an abstract onset and offset emphasis stage. In summary, the MSW model is able to predict the effects of adding temporal fringes on target ILD sensitivity under various temporal configurations. Importantly, the model evaluates the entire AN response to both target and fringes and no windowing of the target response or onset and offset emphasis was required to predict the pattern of results.

4. Hartmann and Constan (2002)

Finally, we applied the model to predict the effect of varying the stimulus' interaural correlation on ILD thresholds reported by Hartmann and Constan (2002). This task indicates the susceptibility of the binaural level comparison stage to interaural differences in the stimulus temporal structure. The stimuli used in their experiment 2 considered here were 500-ms low-pass filtered (1-kHz) white noise tokens. Three binaural conditions were evaluated: interaurally correlated, interaurally anticorrelated, and interaurally uncorrelated. Figure 11 shows the mean measured ILD thresholds ($\pm$ standard deviations) with empty square symbols connected with a solid line. The thresholds were reported to be significantly higher ($p = 0.05$, thus, just at the border of significance) for uncorrelated than for correlated noises, and the anticorrelated noises did not significantly differ from the other two conditions. For the model simulations, ten independent noise tokens were generated for each condition and each ILD. The filled square symbols connected with a dotted line in Fig. 11 show the predictions of the MSW model, qualitatively predicting the threshold elevation from the correlated to the uncorrelated condition. Although the percentage of variance accounted for by the model is negligible (which is due to the overall small effect of interaural correlation in this task), the predictions fall within the error bars of the data. While increasing the window duration tended to improve the prediction accuracy, in line with Brown and Tollin (2016), we considered optimization of the window duration to exceed the scope of the current study.

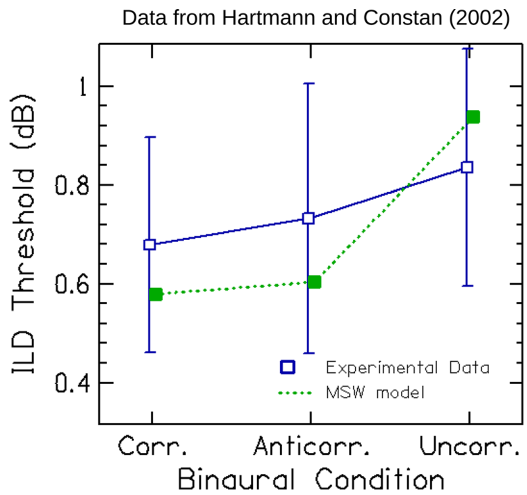

FIG. 11. (Color online) Empty squares connected with solid line: ILD thresholds (± standard deviations) as a function of the interaural correlation of the target stimulus (Corr. = correlated; Anticorr. = anticorrelated; Uncorr. = uncorrelated), adapted from Hartmann and Constan (2002). Filled squares connected with dashed line: Predictions of the MSW model.

C. Discussion

Overall, the model predicted the results from four experimental studies involving various manipulations of temporal stimulus aspects. Given that the model is based on the subtraction of the AN discharge rates between the two ears and interaural timing differences are not explicitly encoded by the model, the auditory system appears to rely mainly on discharge rate differences to extract ILD cues in these tasks. However, one particular condition, namely, pulse trains with a rate of 100 pps, was not correctly predicted. As demonstrated in Fig. 6 (upper right panel), the ongoing model response predicts lateralization to the side with the lower level for this condition, whereas human listeners showed normal lateralization to the side with the higher level in the respective experimental condition. This seems to indicate that listeners use, besides the time-averaged interaural differences in discharge rate, some other cue which is not captured by the model. A potential candidate cue is the latency in the neural response to a stimulus which is known to decrease with increasing stimulus level (Adrian, 1928; Heil and Neubauer, 2001). It has been suggested that the level dependency of latency may convert ILDs to ITDs and the binaural system may actually be sensitive to the ITD cue (Jeffress, 1948). This mechanism appears to be most pronounced at the stimulus onset (Heil, 1998; Irvine *et al.*, 2001; Moller, 1975) and weaker during the ongoing signal (Michelet *et al.*, 2012). For high-frequency sounds as used in the current study, such a latency mechanism would be expected to operate best for stimuli eliciting salient envelope ITD cues. Interestingly, it is well known that envelope ITD sensitivity for AM stimuli is best around 100 Hz and—most relevant in the present context—monotonically worsens toward higher modulation rates (e.g., Bernstein and Trahiotis, 2014, 2002; Majdak and Laback, 2009; Noel and Eddington, 2013). This suggests that for the 100-Hz pulse trains used in the present study human listeners may have relied more on timing cues rather than on discharge-rate-based cues. We surmise that there is a timing component at low pulse rates, which diminishes with increasing pulse rate. Given the onset dominance of the latency effect, an attractive idea is that in low-rate pulse trains each pulse acts as an "onset" so that the latency of the response to each pulse is affected by level, but that this is less the case at higher pulse rates, where only the timing of the response to the first pulse is strongly affected by stimulus level but the subsequent responses are not. This concept is in line with the physiological observations to sustained stimuli quoted above. It is also in line with the rate-dependence of onset ITD sensitivity: Hu *et al.* (2017) argue that up to 200 Hz carrier frequency, each carrier cycle forms a new onset, whereas cycles are fused at higher rates, leading to a different weighting of onset vs ongoing ITD. For high-frequency sounds as used in the current study, such a latency mechanism would be expected to operate best for stimuli eliciting salient envelope ITD cues. To check the plausibility of this scenario, we analyzed archival physiological responses (methods for physiological recordings are described in Joris and Yin, 1998; click stimuli are described in Joris and Yin, 1995, and Oertel *et al.*, 2000). Responses to pulse trains with variable pulse rate were obtained in globular bushy cells in anesthetized cat. These neurons are an important component of the binaural pathways, providing inhibitory inputs to the LSO and medial superior olive (MSO), and there is no reason to think that the results would be substantially different for the excitatory inputs to either MSO or LSO. Pulse trains over a range of ~60 dB (mostly in 5-dB steps) and rates ranging from 60 up to 1500 pps were presented. A correlogram-based analysis of the latency associated with each level (Joris, 2003; Michelet *et al.*, 2012) was used. The coincidence window duration was 100 μs and seven-point triangular smoothing was applied. The left panel of Fig. 12 shows latency vs level functions for one exemplary fiber (CF = 17.5 kHz) with the pulse rate as the parameter. For rates up to 100 Hz, the latency strongly decreases with increasing level, whereas for higher rates latency depends much less on the level, particularly at intermediate levels. The right panel of Fig. 12 shows the slopes of linear regression lines fitted to the functions for three fibers recorded in three cats (CFs of 15.9, 17.2, and 17.5 kHz). All three fibers show a sharp decline in the slope, i.e., a reduction of level-dependent latency, for pulse rates exceeding about 100 pps. These results suggest that the strength of the latency effect and, thus, of a potential envelope ITD cue, is largest up to low modulation rates around 100 pps for which envelope ITD sensitivity is high and for which our discharge-rate-based model underestimates ILD sensitivity. While this supports the idea that listeners in our ILD experiment used "internally" generated ITD cues for the 100-pps condition, the size of the "internal" ITD cue at the measured ILD threshold of the 100-pps pulse train, calculated based on the mean slopes at 100 pps from the right panel of Fig. 12, is only 6 μs at the input to each LSO.[2] This is clearly smaller than psychophysically measured envelope ITD thresholds in humans for comparable stimuli; Majdak and Laback (2009) reported envelope-ITD thresholds of about 60 μs, corresponding to 30-μs input delays (internal ITDs) at each LSO.

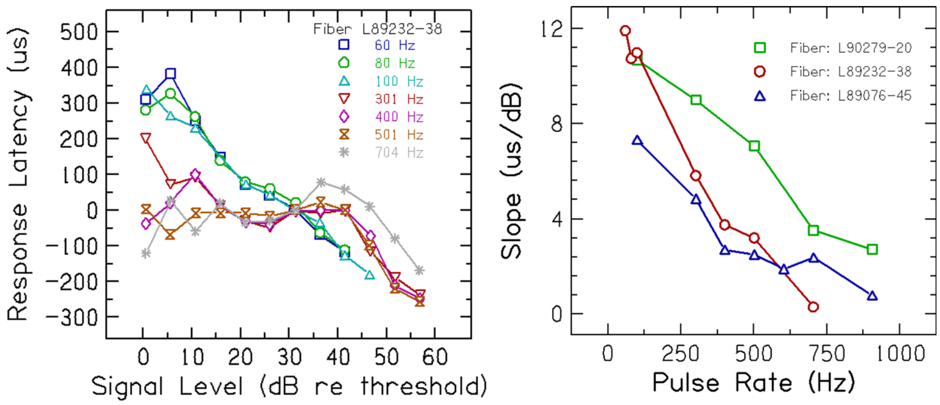

FIG. 12. (Color online) (Left) Response latency (in μs) as a function of signal level (relative to the click train thresholds for the neuron studied) for pulse trains at various rates (see legend) in one exemplary fiber (CF = 17.5 kHz) recorded from the axon of a globular bushy cell in the trapezoid body of an anesthetized cat. The click train threshold for that neuron was 40 dB SPL. (Right) Slopes (in μs/dB) of the response latency vs level functions as a function of pulse rate for three fibers (CFs: 15.9, 17.2, and 17.5 kHz). See Sec. IV C for details.

Other studies measuring at the LSO and using transient stimuli reported larger time-intensity trading ratios that would result in an internal ITD cue at the measured ILD threshold for our 100-pps pulse train condition in the order of 30–60 μs (e.g., Irvine *et al.*, 2001; Park, 1998; Park *et al.*, 1996). This range of internal ITDs is more consistent with the psychophysically measured envelope ITD reported a few lines above. Interestingly, psychophysically measured ITD thresholds were found to be higher at 400 pps than at 200 pps (Majdak and Laback, 2009) and higher at 256 Hz than at 128 Hz (Bernstein and Trahiotis, 2002), consistent with our model-based conclusion that "internally" generated ITD cues contributed in our ILD experiment at 100 pps but to a lesser extent or not at all at 400 pps. We further examined the interaural delay in the response of the AN model by Zilany *et al.* (2014) to 100 - and 400-pps pulse trains at the respective ILD thresholds for those stimuli as measured in the present study. The delay was estimated at the raising flanks of the AN response envelope. Depending on the fiber type, the estimated input delays at each hypothetical LSO amounted to between about 5 and 25 μs for the 100-pps stimuli and between about 5 and 10 μs for the 400-pps stimuli. Thus, for the 100-pps pulse trains the upper limit of input delays estimated from an AN model parametrized for humans roughly agrees with the behavioral ITD sensitivity of humans for that stimulus. Last, we compare behavioral envelope ITD thresholds with ILD thresholds near 100 Hz modulation rate for different envelope shapes: At 128 Hz Dietz *et al.* (2013) reported consistently lower ITD thresholds for transposed tones compared to SAM tones across all levels tested. Based on the latency hypothesis this should result in smaller ILD thresholds as well. A smaller ILD threshold for transposed tones was indeed observed, however, only at 60 dB SPL, which is comparable to the levels tested here.

In summary, it is at least conceivable that human listeners use ITD cues resulting from level-dependent response latency to accomplish threshold ILDs for amplitude modulated sounds if the envelope rate is low (on the order of 100 Hz). However, alternative explanations are conceivable. Ashida *et al.* (2016) recently showed that an LSO model that is based on coincidence detection between ipsilateral excitatory and contralateral inhibitory inputs explains a large variety of data on monaural and binaural AM coding. Such a model may explain at least some of the temporal effects reported here. First, it may explain enhanced ILD sensitivity when imposing AM. For high-frequency pure tones, spiking would be stochastic, leading to low coincidence detection, whereas for pulse trains the envelope phase locking would enhance coincidence detection. Second, it may explain the lower ILD sensitivity for 800-pps compared to 400-pps pulse trains due to the reduction of envelope phase locking in LSO afferents. For 100-pps pulse trains, envelope phase locking would also be strong, but level-dependent response latency might result in temporally misaligned LSO inputs and therefore relatively lower coincidence detection compared to higher rates. It would therefore be interesting for a future study to compare predictions of the coincidence-detection model with predictions of the current discharge-rate-difference model, both combined the same auditory periphery front-end.

Our MSW model accounted for 82% of the variance, on average, in the experimental data of the ILD studies that were modeled (disregarding the interaural correlation experiment, where the variance metric appears not to be useful). The prediction power was slightly lower (73%) when the short-time windowing was omitted (i.e., our LTI model), suggesting that for the experiments under consideration it does not make a great difference if the internal ILD cue is calculated based on the integrated discharge rate and its variation across the entire stimulus duration (LTI model, similar to the loudness meter model by Hartmann and Constan, 2002) or based on short-time windowing (MSW model). While both models have properties enabling them to predict effects of the temporal stimulus structure (e.g., modulation rate) and target duration, the MSW model is more plausible in terms of its window duration, which was chosen to be consistent with the duration of inhibitory effects at the contralateral input to the LSOs for acoustically intact stimuli (e.g., Joris and Yin, 1995). It should be noted, however, that the window duration appears to increase along the ascending auditory pathway (Park, 1998), particularly for temporally degraded stimuli, e.g., as resulting from reverberation (Brown and Tollin, 2016). Thus, longer window durations would likely be required also with the current model to better predict the effects of temporal signal degradation (such as interaural decorrelation).

A somewhat surprising result was that both present model variants account well for the effects of diotic fringes on target ILD discrimination when evaluating the response to the entire stimulus (fringes and target) without specific selection of the target response. We actually also tested if a version of the MSW model in which the windows were selected based on an interaural discharge rate difference criterion would further improve the prediction accuracy (not reported). This approach selected those windows that mainly contained the response to the target (having nonzero ILD) and discarded the windows mainly responding to the fringes (having zero ILD). Unexpectedly, the model performed best when applying no minimum interaural discharge rate criterion at all, thus, including all windows. This result seems to suggest that the listeners in Le Goff *et al.* (2012) did not follow a multiple looks-like approach (Viemeister and Wakefield, 1991), i.e., picking out the "optimal" windows centered on the response to the target and disregarding the windows responding to the fringes. Models following such a strategy by centering a temporal window on the target and thereby attenuating the fringes have been shown to successfully predict experimental data (Akeroyd and Bernstein, 2001; Bernstein *et al.*, 2001; Le Goff *et al.*, 2012). It thus appears that realistic modeling of monaural AN response, as in the current model, already accounts for effects of diotic fringes, suspending the need for the application of a binaural temporal window.

The success of the current model to predict the data by Le Goff *et al.* (2012) also suggests, in line with previous suggestions (e.g., Brown, 2012; Hartung and Trahiotis, 2001), that at least some of the onset (and partially also offset) enhancement often postulated to account for precedence or other temporal weighting effects (e.g., Akeroyd and Bernstein, 2001; Le Goff *et al.*, 2012; Stecker and Hafter, 2009; Stellmack *et al.*, 1999) occurs already at the level of the AN. Importantly, this is not to say that higher-order auditory processing does not also contribute, for example, in context and experience relevant tasks (see Litovsky *et al.*, 1999).

Our model included discharge-rate-based weighting of the individual AN response windows (or bins) to avoid that neural responses before stimulus onset, after stimulus offset, or during silent signal portions disrupt the estimation of the ILD. Although this parameter was not formally optimized because of computational constraints, reasonably good results were obtained for all experiments when setting the value at which the weighting function reaches its maximum (*fmax*) to 32% of the dynamic range of the discharge rates for the particular fiber types used. Note that while further "tuning" of *fmax* was found to result in even better predictions of individual experiments, we preferred to use a common parameter value across experiments.

Despite the roughly similar prediction power of the MSW and LTI model variants, a theoretical advantage of the MSW approach is that it would be better suited to track dynamically changing ILD (e.g., as a result of head or source movement). In order to predict data on dynamic ILDs, future modeling studies would need to extend the MSW model to track dynamic changes of the internal ILD across short-time windows. One challenge will be to distinguish between dynamic changes of the internal ILD due to either changes in the ILD of the external sound source or due to nonlinearities in the front-end model of the auditory periphery, which arise even for a stimulus with constant ILD.

V. SUMMARY AND OVERALL CONCLUSIONS

Temporal effects in ILD-based lateralization discrimination and SLD were studied experimentally. The ILD data and data from a number of published studies on temporal effects in ILD perception were predicted using a model combining an established model of the auditory periphery up to the AN with a relatively simple binaural comparison stage. The main findings are summarized in the following:

- ILD thresholds for constant-duration and high-frequency filtered pulse trains formed a nonmonotonic pattern across pulse rates, with best sensitivity at 400 pps. The increasing sensitivity from 100 to 400 pps is consistent with temporal integration across a larger number of temporal units of ILD information. The reduction in sensitivity from 400 to 800 pps and the further reduction for unmodulated pure tones are consistent with the so-called binaural adaptation effect reported in the literature. The lower ILD sensitivity for pure tones compared to pulse trains suggests that ongoing envelope modulation enhances ILD sensitivity.
- Compared to ILD sensitivity, SLD sensitivity for pulse trains was overall lower (thresholds are about two times higher) and did not degrade from 400 to 800 pps. The overall lower SLD sensitivity may point to a weaker short-term memory for sequential loudness discrimination than for sequential ILD-based lateralization discrimination. The different patterns as a function of rate may suggest different time constants and/or additional processing stages in the coding of loudness as compared to ILD.
- Our model, based on the interaural difference in neural discharge rate, estimated within and integrated across short-time windows, overall well explained present and literature ILD data, including temporal effects with pulse trains, the detection of ILD in a target stimulus surrounded by diotic fringes, and the effect of interaural coherence on ILD sensitivity. Because of the short-time windowing, the model is potentially suited as a framework for predicting the perception of moving sound sources based on dynamic ILDs. Overall, temporal effects in ILD perception for simple and non-degraded stimuli as used in the present study appear to be attributable to known properties of peripheral coding up to the AN, with no need for a particular binaural adaptation mechanism, which has been proposed to operate beyond the level of the AN (e.g., at the cochlear nucleus, as suggested by Hafter, 1997). This indirectly supports the conclusion that SLD perception, showing a different pattern across rates, involves additional processing stages.
- The model systematically underestimated (or not predicted at all) the actual listener's ILD sensitivity for 100-pps pulse trains. We tested the idea that at that rate listeners may use internally generated envelope ITD cues arising from the level dependence in AN response latency. Such ITD cues were not used by the model. The pattern of response latencies across pulse rates obtained from measurements in the trapezoid body of cats are consistent with this idea,

although the absolute size of the latencies was small compared to psychophysical ITD measurements in humans. If latency cues indeed play a role in ILD perception, they may explain also the largely elevated SLD thresholds relative to ILD thresholds at 100 pps.

ACKNOWLEDGMENTS

We thank the two anonymous reviewers for their valuable comments, which helped to improve the quality of an earlier version of the manuscript. We would like to thank Michael Mihocic for implementing parts of the software used for running the experiments and our listeners for their patience during the experiments. We also thank Stephan Ewert for helpful comments on an earlier version of the manuscript. Parts of the work were conducted while the first author was a fellow at the Hanse Institute for Advanced Study (HWK), Delmenhorst, Germany. Funding for that fellowship was provided by the HWK and the Politzer foundation. Funding for M.D. was provided by EU-FP7, ABCIT grant agreement (Project No. 304912). Funding for P.J. was provided by Grant No. 091214N from the Research Foundation Flanders (FWO), Belgium.

[1]The formula used to compute the percentage of the variance for which our predicted thresholds accounted was $100(1 - [\Sigma(E_i - P_i)^2]/[\Sigma(E_i - E\_)^2])$ where E_i and P_i represent individual experimentally measured and predicted threshold values, respectively, and $E\_$ represents the mean of the measured thresholds (e. g., Bernstein and Trahiotis, 2002).

[2]Because the ILD was split between the two intervals of a trial in our experiments, we consider here the latency corresponding to half of the reported threshold ILD.

Adrian, E. (**1928**). *The Basis of Sensation: The Action of the Sense Organs* (Norton, New York), pp. 1–122.

Akeroyd, M. A., and Bernstein, L. R. (**2001**). “The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses,” J. Acoust. Soc. Am. **110**, 2516–2526.

American National Standards Institute (ANSI) (**1996**). *Specifications for Audiometers* (ANSI, New York).

Ashida, G., Kretzberg, J., and Tollin, D. J. (**2016**). “Roles for coincidence detection in coding amplitude-modulated sounds,” PLoS Comput. Biol. **12**, e1004997.

Baumgartner, R., Majdak, P., and Laback, B. (**2016**). “Modeling the effects of sensorineural hearing loss on sound localization in the median plane,” Trends Hear. **20**, 1–11.

Bernstein, L. R. (**2004**). “Sensitivity to interaural intensive disparities: Listeners’ use of potential cues,” J. Acoust. Soc. Am. **115**, 3156–3160.

Bernstein, L. R., and Trahiotis, C. (**2002**). “Enhancing sensitivity to interaural delays at high frequencies by using ‘transposed stimuli,’ ” J. Acoust. Soc. Am. **112**, 1026–1036.

Bernstein, L. R., and Trahiotis, C. (**2014**). “Sensitivity to envelope-based interaural delays at high frequencies: Center frequency affects the envelope rate-limitation,” J. Acoust. Soc. Am. **135**, 808–816.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). “Sensitivity to brief changes of interaural time and interaural intensity,” J. Acoust. Soc. Am. **109**, 1604–1615.

Bos, C., and de Boer, E. (**1966**). “Masking and discrimination,” J. Acoust. Soc. Am. **39**, 708–715.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). “Binaural processing model based on contralateral inhibition. I. Model structure,” J. Acoust. Soc. Am. **110**, 1074–1088.

Brown, A. D. (**2012**). “Temporal weighting of binaural cues for sound localization,” Dissertation, University of Washington, Washington.

Brown, A. D., and Stecker, G. C. (**2010**). “Temporal weighting of interaural time and level differences in high-rate click trains,” J. Acoust. Soc. Am. **128**, 332–341.

Brown, A. D., and Tollin, D. J. (**2016**). “Slow temporal integration enables robust neural coding and perception of a cue to sound source location,” J. Neurosci. **36**, 9908–9921.

Buell, T. N., and Hafter, E. R. (**1988**). “Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times,” J. Acoust. Soc. Am. **84**, 2063–2066.

Buell, T. N., and Hafter, E. R. (**1991**). “Combination of binaural information across frequency bands,” J. Acoust. Soc. Am. **90**, 1894–1900.

Bures, Z., and Marsalek, P. (**2013**). “On the precision of neural computation with interaural level differences in the lateral superior olive,” Brain Res. **1536**, 16–26.

Buus, S. (**1990**). “Level discrimination of frozen and random noise,” J. Acoust. Soc. Am. **87**, 2643–2654.

Devore, S., and Delgutte, B. (**2010**). “Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: Influences of interaural time and level differences,” J. Neurosci. **30**, 7826–7837.

Dietz, M., Bernstein, L. R., Trahiotis, C., Ewert, S. D., and Hohmann, V. (**2013**). “The effect of overall level on sensitivity to interaural differences of time and level at high frequencies,” J. Acoust. Soc. Am. **134**, 494–502.

Glasberg, B. R., and Moore, B. C. J. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1967**). “Auditory nonlinearity,” J. Acoust. Soc. Am. **41**, 676–689.

Grantham, D. W. (**1984**). “Interaural intensity discrimination: Insensitivity at 1000 Hz,” J. Acoust. Soc. Am. **75**, 1191–1194.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Haynes, D. S., and Labadie, R. F. (**2008**). “Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using CIS+ processing,” Ear. Hear. **29**, 33–44.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York), pp. 1–138.

Hafter, E. R. (**1997**). “Binaural adaptation and the effectiveness of a stimulus beyond its onset,” in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 211–232.

Hafter, E. R., and Buell, T. N. (**1990**). “Restarting the adapted binaural system,” J. Acoust. Soc. Am. **88**, 806–812.

Hafter, E. R., and Dye, R. H. J. (**1983**). “Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,” J. Acoust. Soc. Am. **73**, 644–651.

Hafter, E. R., Dye, R. H., Nuetzel, J. M., and Aronow, H. (**1977**). “Difference thresholds for interaural intensity,” J. Acoust. Soc. Am. **61**, 829–834.

Hafter, E. R., Dye, R. H. J., and Wenzel, E. (**1983**). “Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number,” J. Acoust. Soc. Am. **73**, 1708–1713.

Hartmann, W. M., and Constan, Z. A. (**2002**). “Interaural level differences and the level-meter model,” J. Acoust. Soc. Am. **112**, 1037–1045.

Hartung, K., and Trahiotis, C. (**2001**). “Peripheral auditory processing and investigations of the ‘precedence effect’ which utilize successive transient stimuli,” J. Acoust. Soc. Am. **110**, 1505–1513.

Heil, P. (**1998**). “Neuronal coding of interaural transient envelope disparities,” Eur. J. Neurosci. **10**, 2831–2847.

Heil, P., and Neubauer, H. (**2001**). “Temporal integration of sound pressure determines thresholds of auditory-nerve fibers,” J. Neurosci. **21**, 7404–7415.

Houtgast, T., and Plomp, R. (**1968**). “Lateralization threshold of a signal in noise,” J. Acoust. Soc. Am. **44**, 807–812.

Hu, H., Ewert, S. D., McAlpine, D., and Dietz, M. (**2017**). “Differences in the temporal course of interaural time difference sensitivity between acoustic and electric hearing in amplitude modulated stimuli,” J. Acoust. Soc. Am. **141**, 1862–1873.

Irvine, D. R., Park, V. N., and McCormick, L. (**2001**). “Mechanisms underlying the sensitivity of neurons in the lateral superior olive to interaural intensity differences,” J. Neurophysiol. **86**, 2647–2666.

Jeffress, L. A. (**1948**). “A place theory of sound localization,” J. Comp. Physiol. Psychol. **41**, 35–39.

Joris, P. X. (**2003**). “Interaural time sensitivity dominated by cochlea-induced envelope patterns,” J. Neurosci. **23**, 6345–6350.

Joris, P. X., and Yin, T. C. (**1995**). “Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences,” J. Neurophysiol. **73**, 1043–1062.

Joris, P. X., and Yin, T. C. (**1998**). "Envelope coding in the lateral superior olive. III. Comparison with afferent pathways," J. Neurophysiol. **79**, 253–269.
Kidd, G., Jr., Mason, C. R., Best, V., and Marrone, N. (**2010**). "Stimulus factors influencing spatial release from speech-on-speech masking," J. Acoust. Soc. Am. **128**, 1965–1978.
Laback, B., Pok, S.-M., Baumgartner, W.-D., Deutsch, W. A., and Schmid, K. (**2004**). "Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors," Ear. Hear. **25**, 488–500.
Le Goff, N., Kohlrausch, A., and Dau, T. (**2012**). "Effects of diotic fringes on interaural disparity detection (L)," J. Acoust. Soc. Am. **132**, 2959–2962.
Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.
Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.
Majdak, P., Goupell, M. J., and Laback, B. (**2011**). "Two-dimensional localization of virtual sound sources in cochlear-implant listeners," Ear Hear. **32**, 198–208.
Majdak, P., and Laback, B. (**2009**). "Effects of center frequency and rate on the sensitivity to interaural delay in high-frequency click trains," J. Acoust. Soc. Am. **125**, 3903–3913.
Mao, J., and Carney, L. H. (**2015**). "Tone-in-noise detection using envelope cues: Comparison of signal-processing-based and physiological models," J. Assoc. Res. Otolaryngol. **16**, 121–133.
Michelet, P., Kovacic, D., and Joris, P. X. (**2012**). "Ongoing temporal coding of a stochastic stimulus as a function of intensity: Time-intensity trading," J. Neurosci. **32**, 9517–9527.
Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.
Moller, A. R. (**1975**). "Latency of unit responses in cochlear nucleus determined in two different ways," J. Neurophysiol. **38**, 812–821.
Noel, V. A., and Eddington, D. K. (**2013**). "Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time differences," J. Acoust. Soc. Am. **133**, 2314–2328.
Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (**2000**). "Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus," Proc. Natl. Acad. Sci. U.S.A. **97**, 11773–11779.
Park, T. J. (**1998**). "IID sensitivity differs between two principal centers in the interaural intensity difference pathway: The LSO and the IC," J. Neurophysiol. **79**, 2416–2431.
Park, T. J., Grothe, B., Pollak, G. D., Schuller, G., and Koch, U. (**1996**). "Neural delays shape selectivity to interaural intensity differences in the lateral superior olive," J. Neurosci. **16**, 6554–6566.
Park, T. J., Klug, A., Holinstat, M., and Grothe, B. (**2004**). "Interaural level difference processing in the lateral superior olive and the inferior colliculus," J. Neurophysiol. **92**, 289–301.
Pienkowski, M., and Hagerman, B. (**2009**). "Auditory intensity discrimination as a function of level-rove and tone duration in normal-hearing and impaired subjects: The 'mid-level hump' revisited," Hear. Res. **253**, 107–115.
Rakerd, B., and Hartmann, W. M. (**2010**). "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise," J. Acoust. Soc. Am. **128**, 3052–3063.
Reed, D. K., and van de Par, S. (**2015**). "Lateralization of noise bursts in interaurally correlated or uncorrelated background noise using interaural level differences," J. Acoust. Soc. Am. **138**, 2210–2220.
Rice, S. O. (**1945**). "Mathematical analysis of random noise," Bell Syst. Tech. J. **24**, 46–156.
Sakitt, B. (**1973**). "Indices of discriminability," Nature **241**, 133–134.
Sanes, D. H. (**1990**). "An *in vitro* analysis of sound localization mechanisms in the gerbil lateral superior olive," J. Neurosci. **10**, 3494–3506.
Seeber, B. U., and Fastl, H. (**2008**). "Localization cues with bilateral cochlear implants," J. Acoust. Soc. Am. **123**, 1030–1042.
Søndergaard, P., and Majdak, P. (**2013**). "The Auditory Modeling Toolbox," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), pp. 33–56.
Stecker, G. C., and Brown, A. D. (**2010**). "Temporal weighting of binaural cues revealed by detection of dynamic interaural differences in high-rate Gabor click trains," J. Acoust. Soc. Am. **127**, 3092–3103.
Stecker, G. C., and Brown, A. D. (**2012**). "Onset- and offset-specific effects in interaural level difference discrimination," J. Acoust. Soc. Am. **132**, 1573–1580.
Stecker, G. C., and Hafter, E. R. (**2009**). "A recency effect in sound localization?," J. Acoust. Soc. Am. **125**, 3914–3924.
Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (**2005**). "Location coding by opponent neural populations in the auditory cortex," PLoS Biol. **3**, e78.
Stellmack, M. A., Dye, R. H. J., and Guzman, S. J. (**1999**). "Observer weighting of interaural delays in source and echo clicks," J. Acoust. Soc. Am. **105**, 377–387.
Stellmack, M., Viemeister, N. F., and Byrne, A. J. (**2004**). "Monaural and interaural intensity discrimination: Level effects and the 'binaural advantage,' " J. Acoust. Soc. Am. **116**, 1149–1159.
Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.
Takanen, M., Santala, O., and Pulkki, V. (**2014**). "Visualization of functional count-comparison-based binaural auditory model output," Hear. Res. **309**, 147–163.
Tollin, D. J. (**2003**). "The lateral superior olive: A functional role in sound source localization," The Neuroscientist **9**, 127–143.
Tollin, D. J., Koka, K., and Tsai, J. J. (**2008**). "Interaural level difference discrimination thresholds for single neurons in the lateral superior olive," J. Neurosci. **28**, 4848–4860.
Tsai, J. J., Koka, K., and Tollin, D. J. (**2010**). "Varying overall sound intensity to the two ears impacts interaural level difference discrimination thresholds by single neurons in the lateral superior olive," J. Neurophysiol. **103**, 875–886.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.
Xie, B. (**2013**). *Head-Related Transfer Function and Virtual Auditory Display* (J. Ross Publishing, Plantation, FL).
Yost, W. A., and Dye, R. H. J. (**1988**). "Discrimination of interaural differences of level as a function of frequency," J. Acoust. Soc. Am. **83**, 1846–1851.
Zhang, Y., and Wright, B. A. (**2009**). "An influence of amplitude modulation on interaural level difference processing suggested by learning patterns of human adults," J. Acoust. Soc. Am. **126**, 1349–1358.
Zilany, M. S. A., Bruce, I. C., and Carney, L. H. (**2014**). "Updated parameters and expanded simulation options for a model of the auditory periphery," J. Acoust. Soc. Am. **135**, 283–286.
Zilany, M. S. A., Bruce, I. C., Nelson, P. C., and Carney, L. H. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," J. Acoust. Soc. Am. **126**, 2390–2412.