Date of Submission: Fri, May 5, 2017

Psychophysical and modeling approaches towards determining the cochlear phase response based on interaural time differences[(a)]

Hisaaki Tabuchi and Bernhard Laback[(b)]

Austrian Academy of Sciences

Acoustics Research Institute

Wohllebengasse 12-14, 1040 Vienna, Austria

(a) Portions of this work were presented at the 169th Meeting of the Acoustical Society of America, Pittsburgh, 2015, and the 39th MidWinter Meeting of the Association for Research in Otolaryngology, San Diego, 2016.

(b) Bernhard.Laback@oeaw.ac.at

PACS numbers: 43.66.Pn; 43.66.Nm; 43.66.Ba

Keywords: Phase curvature; Phase response; Compression; ROC analysis; ITD

Running Title: Phase response and interaural time differences

ABSTRACT

The cochlear phase response is often estimated by measuring masking of a tonal target by harmonic complexes with various phase curvatures. Maskers yielding most modulated internal envelope representations after passing the cochlear filter are thought to produce minimum masking, with fast-acting cochlear compression as the main contributor to that effect. Thus, in hearing-impaired (HI) listeners, reduced cochlear compression hampers estimation of the phase response using the masking method. This study proposes an alternative approach, based on the effect of the envelope modulation strength on the sensitivity to interaural time differences (ITDs). To evaluate the general approach, ITD thresholds were measured in seven normal-hearing listeners using 300-ms Schroeder-phase harmonic complexes with nine different phase curvatures. ITD thresholds tended to be lowest for phase curvatures roughly similar to those previously shown to produce minimum masking. However, an unexpected ITD threshold peak was consistently observed for a particular negative phase curvature. An auditory nerve response based ITD model, which we first evaluated on published envelope ITD data, predicted the general pattern of our ITD thresholds, except for the threshold peak. Model predictions simulating outer hair cell loss support the feasibility of the ITD-based approach to estimate the phase response in HI listeners.

I. INTRODUCTION

The phase response of a transmission system such as the cochlea determines the relative timing at which different input frequencies are transmitted. Considering a natural sound consisting of multiple frequency components processed by an auditory filter (AF) with a certain characteristic frequency (CF), the temporal output pattern of the AF depends on the phase relations of the sound's spectral components in relation to the AF's phase response. This is particularly the case for higher CFs where multiple spectral components pass the AF. Therefore, the AF phase response, or more generally, the cochlear phase response, is important for encoding temporal cues such as periodicity pitch and interaural time differences (ITDs). The present study proposes and evaluates a method to determine the cochlear phase response that is based on the perception of ITD in multi-tone stimuli. Compared to existing methods, the new approach does not rely on cochlear compression and is, thus, potentially better suited to determine the phase response in listeners with cochlear hearing loss.

An established method to measure the cochlear phase response determines the amount of masking by so-called Schroeder-phase harmonic complexes (SPHCs) with various phase curvatures on a pure tone target (Smith et al., 1986). Importantly, varying the phase curvature of SPHCs does not affect their long-term power spectrum, allowing to independently study the effect of their phase curvature which results in different forms of envelope modulation (as shown in Fig. 1). For a given phase response of the AF centered on the SPHC, variable phase curvatures of the SPHC elicit different amounts of peakedness in their "internal" temporal representations (e.g., after auditory filtering). The rationale of the masking method, as used in some of the later studies (e.g., Lentz and Leek, 2001; Oxenham and Dau, 2004; Tabuchi et al., 2016), is that the SPHC causes a minimal amount of masking if its internal representation is maximally peaked, which is the case if its phase curvature corresponds to the inverse of the AF's phase response. Thus, assuming a uniform phase curvature of the AF, the uniform phase curvature of an SPHC

that yields minimum masked threshold serves as a measure of the AF's phase response.

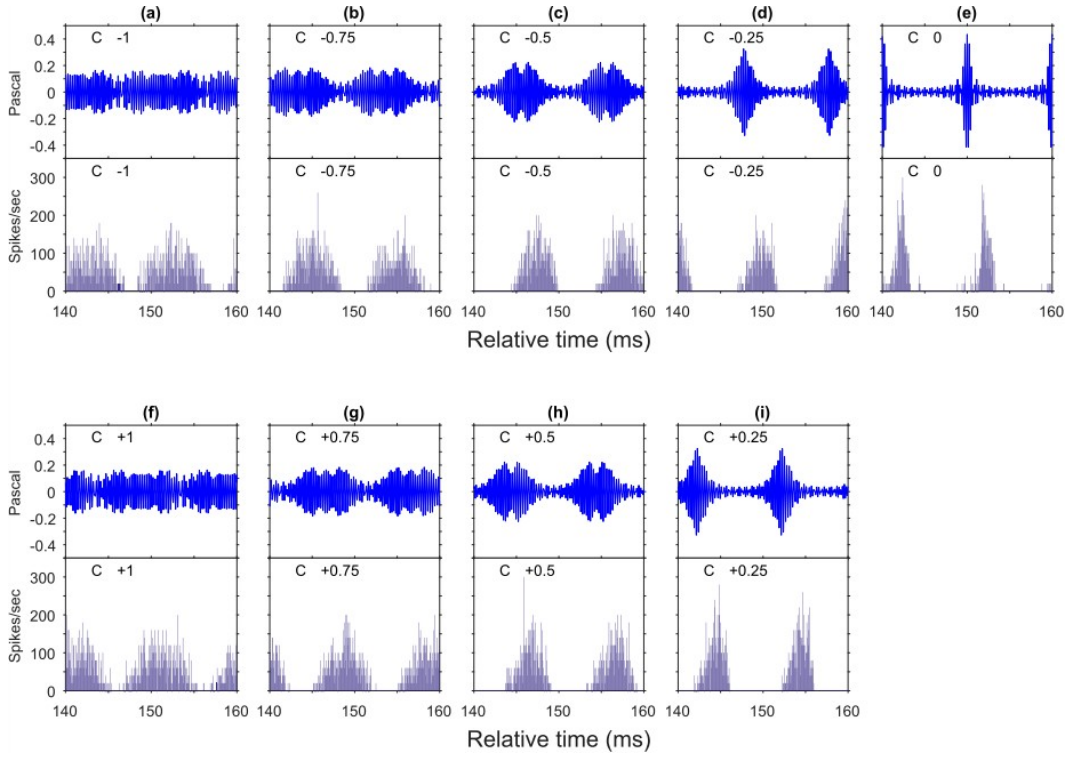

Fig. 1: (Color online) Two-period excerpts of stimulus waveforms and corresponding PSTHs of the AN front-end model (Zilany et al. 2014). (a): C = -1, (b): C = -0.75, (c): C = -0.5, (d): C = -0.25, (e): C = 0, (f): C = +1, (g): C = +0.75, (h): C = +0.5, (i): C = +0.25. The long-term power spectra of the waveforms are constant across Cs.

There are several indications that fast-acting compressive amplification by the outer hair cells (OHCs) in the cochlea is important for the masker-phase effect to occur (e.g., Carlyon and Datta, 1997; Oxenham and Dau, 2004; Wojtczak and Oxenham, 2009; Tabuchi et al., 2016). The basic idea is that maskers eliciting a more peaked internal representation produce a smaller internal excitation level following fast-acting compression, resulting in less masking compared to maskers eliciting a flat representation. Perhaps the strongest indication for the importance of compression is the finding of a large masker-phase effect in forward masking, where alternative explanations for a simultaneous masking configuration, such as detecting the target in the temporal dips of a modulated masker, can be ruled out (Carlyon and Datta, 1997). Second, a recent study showed a reduction of the simultaneous masker-phase effect when adding a

precursor signal supposed to reduce cochlear compression by means of activation of the efferent system (Tabuchi et al., 2016). Third, listeners with cochlear hearing loss, characterized by reduced or absent OHC compression, have been shown to elicit very little or no masker-phase effect (Summers and Leek, 1998; Summers, 2000, 2001; Oxenham and Dau, 2004).

In this study, we propose an alternative approach to determine the cochlear phase response that does not rely on cochlear compression, and is, thus, potentially applicable in hearing-impaired (HI) listeners. The method exploits the listener's sensitivity to ITD of SPHC targets measured with various phase curvatures, and assumes that the peakedness of the internal target representation directly impacts ITD sensitivity. Analogous to the masking method, the SPHC phase curvature associated with the lowest ITD threshold is considered as a measure of the cochlear phase response. The method's assumption of a dependence of ITD sensitivity on the peakedness of the signal's internal envelope representation is based on several recent studies in the field of binaural hearing.

The basis for these binaural studies is the general finding that listeners are sensitive to ITD in the ongoing envelope of sounds (Henning, 1974; Yost, 1976; Bernstein and Trahiotis, 1994) whose carrier frequencies are beyond the so-called fine-structure perception limit of about 1300 Hz (Zwislocki and Feldman, 1956). Bernstein and Trahiotis (2009), Klein-Hennig et al. (2011), Laback et al. (2011), Francart et al. (2012), and Dietz et al. (2015) went on to show that envelope ITD thresholds systematically decrease with increasing peakedness in the ongoing envelope shape, as studied by independent variation of the modulation depth or more detailed parameters of the ongoing envelope such as the flank slope and the pause time within the modulation cycle.

In all of the cited studies on envelope ITD perception, the stimulus envelopes were controlled by temporally shaping a high-frequency pure tone carrier, a process confounded to some degree with spectral alteration. In the present study, we controlled the stimuli by varying

their phase properties, thus, ruling out any potentially confounding changes in the long-term power spectrum. Varying the phase curvature of SPHCs causes variable degrees of envelope peakedness, as shown in Fig. 1. In view of existing masking data at the CF region considered in our study, we hypothesized a V-shaped pattern of ITD thresholds, with lowest thresholds for slightly positive phase curvatures conveying maximally peaked internal envelopes, and elevated thresholds for more positive or negative phase curvatures conveying flatter internal envelopes. Thus, according to our assumptions, the general pattern of ITD thresholds as a function of phase curvature should be roughly similar to the pattern of masked thresholds obtained for similar stimuli, with the minimum representing in both cases a measure of the inverted phase response of the corresponding AF.

An important question is to what extent recently developed computational models are able to predict the effects of varying the stimulus phase curvature on ITD thresholds. Importantly, to correctly predict such data, the model requires a realistic representation of the auditory periphery, including the cochlear phase response. Figure 1 shows the stationary part of SPHC signals used in the present study and the corresponding peri-stimulus time histograms (PSTHs) calculated by a well-established auditory-nerve (AN) model (Zilany et al. 2014). The PSTHs for the different phase conditions generally resemble the temporal shape of the signal envelopes, which implies that the AN front-end model has some sensitivity to the modulation strength of stimuli. Envelope ITD thresholds for several envelope manipulations have been shown to be predictable using a simplified representation of the auditory periphery at the two ears, consisting of a linear Gammatone filter bank, square-law rectification, power-law compression, and envelope low-pass filtering, followed by an interaural comparison based on the normalized cross-correlation (NCC) metric (referred here to as NCC model, e.g., Bernstein and Trahiotis, 2009; Klein-Hennig et al., 2011). However, some effects of envelope variations on ITD perception, such as temporally inverting stimuli with temporally asymmetric envelopes

(e.g., steep-raising and shallow-decaying flanks) or varying the pause duration, were not fully predicted by that model (see Klein-Hennig et al., 2011, and Sec. IV.C.4 of the present paper). Furthermore, the Gammatone filter bank has been shown to not appropriately account for monaural phase effects (Kohlrausch and Sander, 1995; Oxenham and Dau, 2001). We therefore devised a model combining an established nonlinear model of the auditory periphery, potentially exhibiting a more realistic phase response, with a probabilistic interaural comparison stage. To evaluate the model's general capability in predicting envelope ITD thresholds, it was also tested on literature data on systematic variation of envelope shape parameters.

In this study we performed an experiment with seven normal-hearing (NH) listeners on the effect of systematically varying the stimulus phase curvature on ITD thresholds (Sec. II) and a follow-up experiment with two of those listeners using a finer sampling of phase curvatures (Sec. III). We then evaluated the ability of our model to predict our experimental data and relevant data from the literature (Sec. IV). Finally, we also used the model to predict how OHC loss may affect phase effects (Sec. IV.C.6).

II. EXPERIMENT 1: ITD THRESHOLDS AS A FUNCTION OF PHASE CURVATURE

A. Listeners and equipment

Seven subjects aged between 19 and 33 years participated in the experiment and received monetary compensation for their participation. All had absolute thresholds of 20 dB hearing level or lower at octave frequencies between 0.25 and 8 kHz. Three of the listeners had experience from previous experiments on ITD perception. None of the authors participated in the experiment. All experiments met the ethical principles of the Acoustical Society of America.

The stimuli were generated on a computer and output via a sound interface (E-Mu 0404, Creative Professional) at a sampling rate of 96 kHz and a resolution of 24 bits. The analog

signal was sent through a headphone amplifier (G93, Lake People) to circumaural headphones (HDA 200, Sennheiser). The stimuli were calibrated using an artificial ear (4153, Bruel & Kjær) and a sound level meter (2260, Bruel & Kjær). The experiment was performed in a double-walled sound booth.

B. Stimuli

The SPHCs (Schroeder, 1970; Lentz and Leek, 2001) were defined as:

$$m(t) = \sum_{n=N_1}^{N_2} \cos\left[2\pi n f_0 t + \frac{C\pi n(n+1)}{N_2 - N_1 + 1}\right] \quad (1)$$

where C determines the constant phase curvature, f_o the fundamental frequency, and N_1 and N_2 the lowest and highest harmonics, respectively. The complex had a fundamental frequency of 100 Hz and the harmonics ranged from 3400 to 4600 Hz ($N_1 = 34$, $N_2 = 46$). This stimulus bandwidth is restricted compared to most published masking studies using SPHCs. One reason was to avoid the potentially confounding influence of off-frequency components half an octave or more below CF having zero phase curvature (e.g., Shera, 2001; Oxenham and Ewert, 2005; Tabuchi et al., 2016). A different AF curvature within the off- and on-frequency region would violate the assumption of uniform phase curvature, as imposed by the constant curvature of the stimuli, and likely result in less pronounced phase effects. A second reason was to reduce the possibility that listeners use multiple AFs, including those remote from the target center frequency, to extract and combine ITD cues. A third reason was that all harmonics should be sufficiently above the fine-structure sensitivity limit of human listeners to avoid the contribution of fine-structure ITD cues (e.g., Zwislocki and Feldman, 1956; Brughera et al., 2013), and thus potentially dominate ITD sensitivity. The stimuli had a total duration of 300 ms including 125-ms cosine-squared on- and off ramps to minimize onset and offset ITD cues. They were presented at an overall sound pressure level (SPL) of 70 dB (re 20 µPa). Nine *C*s ranging from -1

to 1 in intervals of 0.25 were tested. Figure 1 shows example waveforms of the stimuli for *C*s from -1 to +1. Although the long-term power spectra (across envelope periods) are constant across *C*s (see Kohlrausch and Sander, 1995), the stimuli clearly differ in the peakedness of their envelopes, being largest for $C = 0$ and decreasing towards $C = \pm 1$. The ITD cue was applied to the target stimulus by delaying the entire waveform at one ear.

Interaurally uncorrelated background noise was continuously presented at 55 dB SPL to mask low-frequency cochlear distortion products. The background noises on the two ears were generated by low-pass filtering Gaussian white noises with a second-order Butterworth filter at a cut-off frequency of 1300-Hz and with an attenuation of 12 dB/oct.

C. Procedure

Percent-correct scores in a left/right discrimination task were measured as a function of ITDs (50, 100, 200, 400, 800, and 1600 μs). The first interval in a trial always contained the reference stimulus with *zero* ITD and the second interval contained the target stimulus with a *non-zero* ITD. The listeners indicated if the target stimulus was to the left or right of the reference stimulus using a response pad. The target ITD had equal a priori probability of leading at the left and right ear. Each stimulus interval was signaled visually on a computer screen, and the between-interval gap was 300 ms. Feedback on the correctness of the response was provided visually after each trial.

A block consisted of 540 presentations of nine *C*s and six ITDs, with ten repetitions of these conditions in randomized order. Each listener completed six blocks. Thus, each combination of *C* and ITD was tested 60 times. The total testing time of the experiment amounted to about six hours per listener. We checked the ITD thresholds of each listener across blocks, but observed no systematic learning effects.

Based on the psychometric functions for each *C* and listener, ITD thresholds at the 80-

%-point of the psychometric function were estimated using the maximum-likelihood method in combination with a two-parameter Weibull function fit (Myung, 2003). The grand mean thresholds were calculated as the geometric means and standard deviations of ITD thresholds across individual listeners.

Before commencing the main experiment, the listeners completed a training session using the same procedure as in the main experiment. The stimulus was a bandpass-filtered white noise spectrally centered at 4600 Hz and with a bandwidth of 1500 Hz. Blocks of 100 trials with a fixed ITD value (100, 200, 400, or 600 μs) were run. The training started with the 600-μs ITD and continued towards smaller ITDs until the performance for all the ITD values became better than 80 % correct.

D. Results and discussion

The psychometric functions were found to monotonically increase across ITDs and the Weibull function yielded reasonable fits to each of the psychometric functions. The amount of variance explained by the Weibull fit was quantified for each *C* and listener as a percentage, with the average across all the *C*s and listeners amounting to 89.1 % (standard deviation: 10.2 %). Fig. 2 shows the ITD thresholds as a function of *C* for the individual listeners and the across-listener means (error bars showing ±1 standard deviation). In some listeners, it appears difficult to determine a minimum threshold; for example, the thresholds of $C = 0$ and 0.5 in NH39, and $C =$ -0.75, -0.25, and 0 in NH47, are close to each other. Despite differences across individual listeners, the across-listener means reveal a pattern of low thresholds around $C = 0$ and increasing thresholds towards $C = \pm 1$, with the exception of an elevation at $C = -0.5$ (see below). The overall pattern of ITD thresholds is consistent with previous studies in showing that the ITD sensitivity changes with the peakedness of the envelope shape (e.g., Bernstein and Trahiotis, 2009; Klein-Hennig et al., 2011; Laback et al., 2011). We are not aware of published data on the

effects of direct manipulation of the stimulus phase properties on ITD sensitivity with which the current data could be compared with. A one-way repeated-measures ANOVA indicated a significant effect of phase curvature C [$F(8, 48)=7.25$, $p<0.001$]. Post-hoc pairwise comparisons using the Tukey LSD test indicated that thresholds do not differ significantly between -0.75, -0.25, 0, 0.25, and 0.5 ($p>0.07$), but thresholds differ significantly between $C = 0$ versus $C = -1$, -0.5, 0.75, and 1 ($p<0.03$). The phase curvatures showing the lowest mean ITD threshold (around $C = 0$) are roughly consistent with those that produced minimum masking in a study (Tabuchi et al., 2016) using similar stimuli in six of the listeners of the present study (between $C = 0.25$ and $C = 0.5$). For comparison, the Cs showing minimum masked thresholds for those listeners are indicated by small arrows in Fig. 2. The threshold minima are clearly more variable and less pronounced for the ITD paradigm than for the masking paradigm (see Tabuchi et al, 2016). This suggests that the ITD thresholds are less stable and reliable than the masked thresholds, an issue that is addressed in Experiment 2.

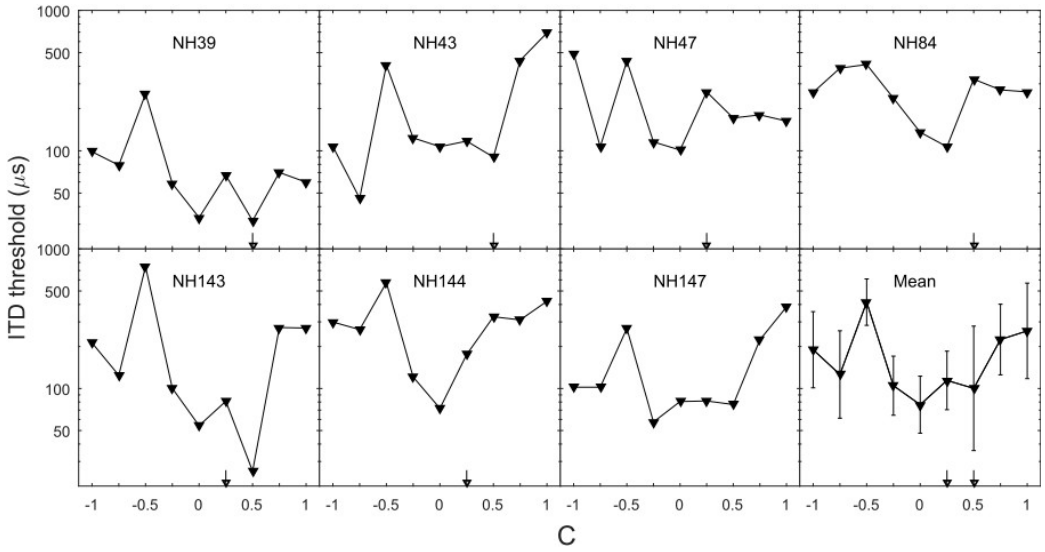

Fig. 2: (Color online). Results of Experiment 1: ITD thresholds are plotted as a function of C for individual listeners. Mean ITD thresholds across listeners (±1 standard deviation) are shown in the bottom right panel. The small arrows indicate the Cs exhibiting minimum masked thresholds in a masking experiment using similar stimulus conditions with the respective listeners (Tabuchi et al., 2016).

An unexpected non-monotonic threshold elevation (peak) was consistently observed at C = -0.5 for all listeners. The post-hoc comparisons indicated that this threshold peak significantly differed from the thresholds at surrounding Cs (-0.25 and -0.75) and even from C = -1 (p<0.019). Inspection of the stimulus envelope shapes for the stimuli with various Cs (Fig. 1) did not reveal any convincing explanation for the threshold peak in terms of envelope peakedness. In particular, the stimulus with C = -0.5 is not less peaked than the stimuli with Cs of -0.75 and -1, as would be expected by its higher threshold. Because the "internal" stimulus representation actually depends on the interaction between the stimulus phase and the phase response of the cochlea, we will address this issue in more detail in the modeling section below (Section IV).

III. EXPERIMENT 2: ITD THRESHOLDS AT FINER STEPS OF NEGATIVE PHASE CURVATURES

This experiment served to sample the thresholds at finer steps of C in the vicinity of the threshold peak at C = -0.5 found in Experiment 1 and check the reproducibility of the threshold peak. Two listeners (NH143 and NH144) participated in Experiment 2, almost one year after they completed Experiment 1. The equipment, stimuli, and procedure were all the same as those described in Sec. II, except that the stimuli consisted of several additional Cs and that positive Cs were not tested. The following Cs were tested: -1, -0.75, -0.67, -0.59, -0.5, -0.42, -0.34, -0.25, and 0.

The thresholds of Experiment 2 were estimated from psychometric functions by using a Markov chain Monte Carlo technique for Bayesian inference which has been recently developed in vision research (Fründ et al., 2011). 95% confidence intervals of threshold estimates were calculated using a bootstrap technique[1]. We observed that the Markov chain Monte Carlo technique and the maximum likelihood method (Myung, 2003) used in Experiment

1 resulted in very similar threshold estimates.

A. Results and discussion

Fig. 3a and 3b show the individual listener's thresholds as compared to the corresponding individual's thresholds from Experiment 1 (the latter replicated from Fig. 2). The error bars show 95-% confidence intervals of the threshold estimates. For listener NH143, the thresholds for phase curvatures measured both in Experiment 1 and 2 show high reproducibility, including the threshold peak at C = -0.5, as can be seen from the large overlap of the 95-% confidence intervals. That listener's data from Experiment 2 for phase curvatures not tested in Experiment 1 show an additional sharp peak at C = -0.34. The data of listener NH144 generally show much more pronounced differences across the two experiments, although the overall pattern, including the peak at C = -0.5, is preserved. That second listener's data show no additional threshold peak. In summary, these data suggest some uncertainty in the ITD-threshold estimates, at least for individual listeners. Regarding the non-monotonicities in the threshold functions, the origin is unclear at this point. Modeling analyses are provided in Sec. IV.C.2 and IV.C.3 to explore possible explanations.

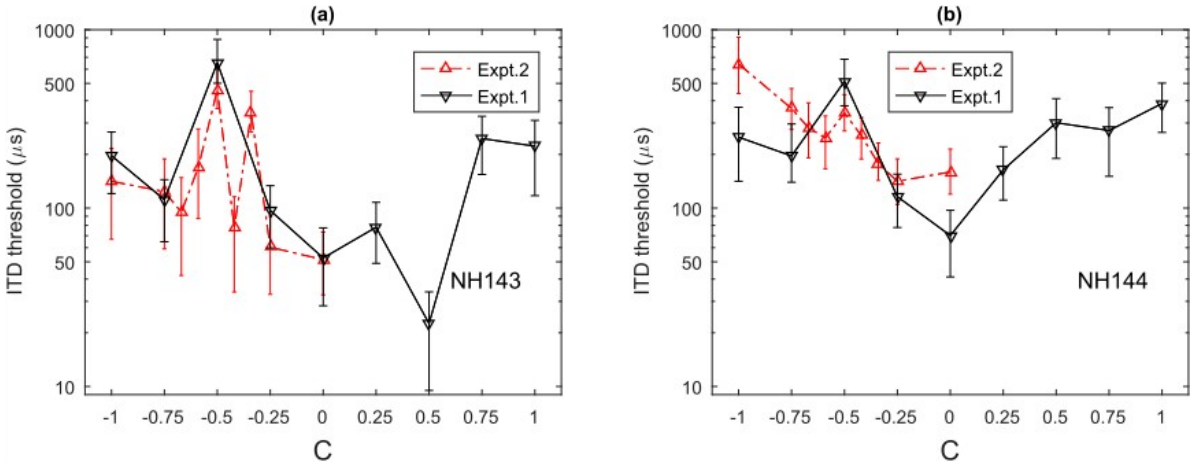

Fig. 3: (Color online) Results of Experiment 2: ITD thresholds for finer C steps (upward pointing triangles) measured in individual listeners. (a): NH143, (b): NH144. The thresholds for these listeners from Experiment 1 are denoted by the downward-pointing triangles. The error bars denote 95-% confidence intervals of the threshold estimates (see text).

IV. MODEL

A. Rationale and general properties

In order to obtain more insight into the mechanisms underlying the experimental results, we performed a modeling approach. As a model of the auditory periphery up to the level of the AN, we used the well-established cat AN fiber model as described in Zilany et al. (2014) with parameters adjusted for humans. This represents the latest version (5.2) of a family of phenomenological models of the transformation of acoustic stimuli into AN discharges, including outer- and middle-ear filtering, cochlear processing, inner hair conductance as well as AN transmission properties. It has been shown to account for many effects of peripheral processing, including level dependent shifts in best frequency, suppression, and AN adaptation. Even more important for the present purpose, the model appears to provide a quite realistic auditory representation of temporal signal aspects, e.g., temporal modulation coding (Zilany et al., 2009) and the phase transfer function (Carney, 1993; Zhang et al., 2001; Zilany and Bruce, 2006). An attractive feature of the AN front-end is its capability of optionally simulating the effects of OHC loss associated with cochlear hearing impairment (see Zilany and Bruce, 2006). The OHC gain is controlled by changing the output of the control path (Bruce et al., 2003). The simulation of complete OHC loss in the model has been shown to be associated with gain reduction in the order of 30 dB or more (resulting in increased absolute thresholds), reduced compression of the input-output function, and reduced frequency selectivity at corresponding CFs (Heinz et al., 2001; Bruce et al., 2003).

Varying the OHC gain allowed us to use the model to predict ITD thresholds across *C* in simulated cochlear hearing impairment associated with OHC loss. Note the implicit assumption of this approach that cochlear hearing loss has no consequences on auditory processing beyond the AN. Moreover, a recent animal study suggests that noise-induced hearing loss induces the selective loss of AN fibers with low- and middle spontaneous firing rates (SRs,

313 Furman et al., 2013). Providing modeling results with different SRs appears therefore useful for
314 understanding the behavioral consequence of OHC loss for specific fiber types.

315 In order to predict behavioral ITD thresholds, we combined two versions of the
316 monaural AN front-end model with a binaural comparison stage. One well-established model of
317 envelope-ITD perception is the NCC model. Although the NCC model has been shown to predict
318 a variety of envelope ITD data (e.g., Bernstein and Trahiotis, 2009), Klein-Hennig et al. (2011)
319 showed that the effects of some envelope manipulations cannot be predicted by that model; for
320 example, temporally inverting an amplitude modulated signal with steep raising flanks but flat
321 decaying flanks in each cycle results in an elevation of ITD thresholds, whereas the NCC model
322 predicts constant thresholds. Adding adaptation loops in the NCC model, intended to replicate
323 some aspects of neural adaptation (Dau et al., 1996), showed somewhat improved threshold
324 predictions for some conditions, but systematic discrepancies from experimental thresholds
325 remained (Klein-Hennig et al., 2011). The front-end of the current model appears to have the
326 potential to even more accurately represent peripheral temporal coding properties.

327 For the binaural stage, we decided for a statistically motivated receiver-operating
328 characteristic (ROC) based approach, mainly because the detailed properties of binaural cue
329 extraction, from early stages such as the medial and lateral superior olives up to the inferior
330 colliculus and the auditory cortex, are still matter of debate and far from being fully understood
331 (e.g., Grothe et al., 2010). Our approach has no *a priori* assumptions on the binaural processing
332 characteristics and can be considered as a behavioral, optimum-observer-like approach that is
333 based on the monaural AN outputs. It thus differs from more sophisticated physiology-based
334 binaural processing stages (e.g., Wang et al., 2014; Gai et al., 2014; Dietz et al., 2016).

335 Our model simulations were based on the left- and right-ear PSTHs of the AN model
336 response to a stimulus, including additional 20 percent of the stimulus duration to account for
337 response delay and decay. Each PSTH was based on 1000 spike train realizations. All monaural

simulations were run 10 times. The binaural simulations were run 10 times in case of normal OHCs and 60 times in case of OHC loss. A larger number of repetitions was chosen for OHC loss because we observed larger variance in the response as compared to normal OHCs. The reported monaural and binaural predictions are based on averaging across those repetitions. We verified for a subset of conditions that the predictions for normal OHC were the same when running either 10 or 60 repetitions, apart from the error bars becoming smaller. The model parameters of interest were the SR of AN fibers (low, medium, or high) and the OHC scaling factor (either 1 = normal or 0 = fully lost). Unless otherwise stated, the following model parameters were fixed: CF: 4 kHz; model sampling rate: 100 kHz; inner hair cell (IHC) scaling factor: 1 (normal); species: human, using basilar membrane tuning from Shera et al. (2002); fractional Gaussian noise type: variable; power-law implementation: approximate; bin width of the PSTH: 50 μs. The input stimuli used in our modeling approach were the same as those used in Experiment 1.

In the following subsection we describe the properties of the AN front-end model in monaurally processing our experimental stimuli. Then, we describe the binaural comparison stage and report on the ability of the complete (binaural) model to predict experimental ITD thresholds.

B. Monaural model analysis

1. Synchronization index (SI) of fundamental frequency for normal and lost OHCs

To study the temporal properties of the AN front-end model we followed the basic idea of temporal synchronization analysis (Goldberg and Brown, 1969). To measure the synchronization of the neural response to the stimulus' temporal envelope modulation, period histograms (PHs) were first obtained by adding up the firing rates (spikes/sec) across the cycles of the PSTHs and

dividing the sum by the total number of firing rates, resulting in the probability density function (the more precise term in a discrete context actually being probability mass function) as shown in Fig. 4. We then computed the SI by calculating the magnitude spectrum of the probability density function as similarly done by Johnson (1980) based on the assumption that the SI is independent of the average firing rate in the analysis. The fundamental frequency (100 Hz) was the spectral component of interest for the analysis because this was the most prominent modulation frequency.

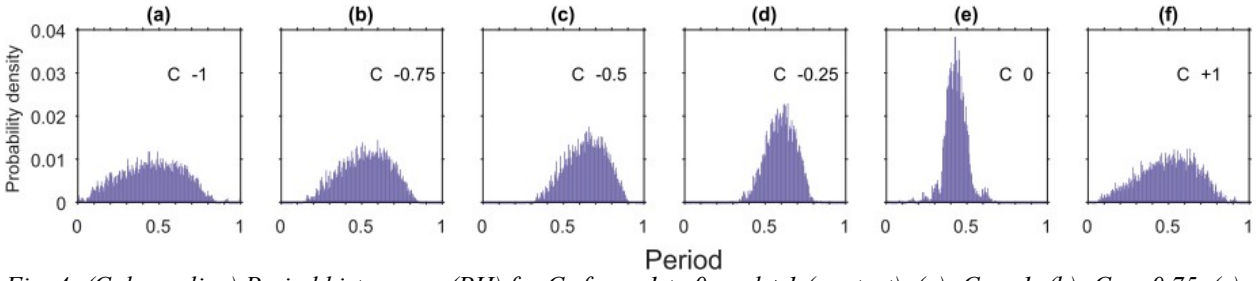

Fig. 4: (Color online) Period histograms (PH) for Cs from -1 to 0, and +1 (see text). (a): C = -1, (b): C = -0.75, (c): C = -0.5, (d): C = -0.25, (e): C = 0, (f): C = +1.

Fig. 5a shows the SI of the fundamental frequency for normal OHCs with different SR fiber types as the parameter. The SI is largest at $C = 0$ and decreases towards $C = \pm 1$. The SI is consistent with the general pattern of ITD thresholds across C reported in Experiment 1, although it does not account for the threshold peak at $C = -0.5$. We will reconsider the threshold peak in the binaural modeling. Overall, apart from the threshold peak, the SI analysis is consistent with Zilany et al. (2009), demonstrating that the AN model robustly predicts the effect of varying the depth of sinusoidal amplitude-modulation (AM) on physiological measures of neural envelope synchrony. It is also evident that the SIs of high-SR fibers are much lower than those of low- and medium-SR fibers. This response pattern across SR fibers is also consistent with physiological data for AM tones, which suggested that low-SR fibers are generally better phase-locked to the modulation frequency than high-SR fibers, especially for CFs below 5 kHz

(Joris and Yin, 1992).

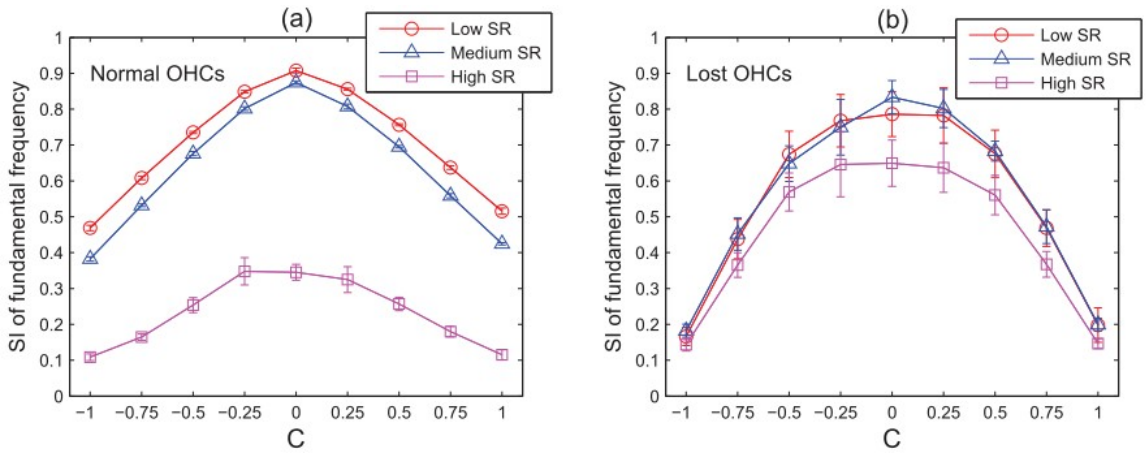

Fig. 5: (Color online) Predicted synchronization index (SI) based on the AN front-end model as a function of C. Symbols indicate the means (±1 standard deviation) across repeated calculations for the three model fiber types with different spontaneous rates (SR). (a): Normal OHCs, (b): OHC loss.

Fig. 5b shows the SI in case of lost OHCs for the three SR fiber types. The general patterns are similar to those for normal OHCs, but there are some systematic differences. First, the SIs for the low- and medium-SR fibers are almost unchanged by OHC loss for $C = 0$, but they are lowered when C approaches ±1. Second, the SI for the high-SR fibers is considerably larger in the vicinity of $C = 0$, while there is almost no difference for $C = \pm 1$. It seems that high-SR fibers with normal OHCs are saturated for the particular stimulus, whereas OHC loss shifts the dynamic range of those fibers to better encompass the dynamic range of the stimulus' temporal envelope; and therefore, results in enhanced temporal coding. This idea was confirmed by additional simulations on SPHCs with a reduced stimulus level (50 dB SPL) in normal OHCs (not shown), revealing enhanced temporal coding (in terms of the SI) for high-SR fibers. The finding is also consistent with the notion that the gain reduction associated with impaired OHC loss enhances neural phase locking (Kale and Heinz, 2010; Henry and Heinz, 2013). In summary, the general finding of increased difference between minimum and maximum SI in simulated OHC loss suggests that phase effects might be stronger in impaired ears compared to normal

ears.

2. Firing rate for normal and lost OHCs

Fig. 6a and 6b show the mean firing rate as a function of *C* with the three SR fibers for normal and lost OHCs, respectively. The patterns of firing rates across *C* are inverted compared to the patterns of SIs from Fig. 5, i.e., they show lowest rates for $C = 0$. These general patterns are consistent with the idea that instantaneous cochlear compression results in lower excitation levels and, thus, lower firing rates for more peaked stimuli (Carlyon and Datta, 1997). In this respect, they are also roughly consistent with a recent masking study using the same stimuli as maskers (Tabuchi et al., 2016), showing less masking for more peaked masker waveforms, although the minimum occurred for *C*s between 0.25 and 0.5. Compared to the predictions for normal OHC function (Fig. 6a), the prediction for OHC loss (Fig. 6b) shows similar differences in the firing rates across *C* in case of high-SR fibers, but largely reduced differences in case of low- and medium-SR fibers. An overall reduction of firing rates as a result of OHC loss is observed, especially for the low- and medium-SR fibers, although these fibers have been shown to well encode stimulus phase differences (see Fig. 5b).

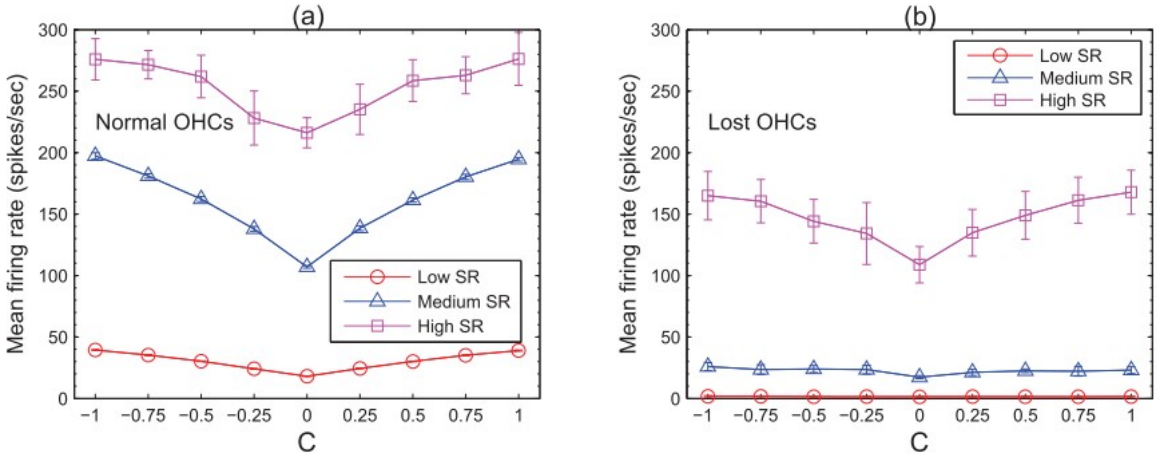

Fig. 6: (Color online) Predicted firing rate based on the AN front-end model as a function of C. Symbols indicate the means (±1 standard deviation) across repeated calculations for the three model fiber types with different spontaneous rates (SR). (a): Normal OHCs, (b): OHC loss. Because the firing rates for low-SR fibers with OHC loss appear essentially at zero in the panel (b), the exact values (in spikes/secs) for the different Cs are denoted in the following: 1.8 (C = -1), 1.7 (C = -0.75), 1.5 (C = -0.5), 1.6 (C = -0.25), 1.4 (C = 0), 1.6 (C = +0.25), 1.5 (C = +0.5), 1.4 (C = +0.75), and 1.6 (C = +1).

C. Binaural model analysis

In the previous section, it was shown that the monaural SI analysis is consistent with the minimum ITD-threshold from Experiment 1 at $C = 0$. In order to more directly predict ITD thresholds, we added a binaural processing stage which is based on minimum assumptions on the physiological mechanism. In order to predict psychophysical ITD thresholds, we compared responses of the front-end AN model to the left and right ear stimuli using the concept of ROC analysis. The left/right discrimination data underlying our ITD thresholds represent a binary classification problem which can be treated with ROC analysis. In analogy to ROC analysis used in psychophysics to describe the observer's sensitivity in a force-choice signal detection task (Green and Swets, 1974), our ROC analysis classified a given relative timing difference between the monaural AN representations, i.e., an ITD, which allowed to predict the sensitivity for that ITD. ITD thresholds were then estimated from predicted neural sensitivity estimates for different ITDs, analogous to psychophysical ITD thresholds estimated from psychometric functions. To illustrate the concept, imagine the simple case of a very short (impulsive) sound presented to a

listener with a given ITD. The neural spikes in response to such a sound would be phase locked to the stimulus and, thus, the left- and right-ear PSTHs would be temporally delayed relative to each other according to the stimulus ITD. The PHs obtained from these PSTHs would likely have Gaussian-like shapes (see e.g. Dreyer and Delgutte, 2006) and fulfill the requirements of ROC analysis to classify the ITD between them (note that a strict assumption of normality is not necessarily required for ROC analysis, as long as the underlying probability density functions decay towards both sides; see, e.g., Hanley and McNeil, 1982). Because our SPHC stimuli have a periodic temporal envelope, for which, by definition, the PH is bounded by the envelope period, the question arises if the PH's shape fulfills the requirement of ROC analysis. Inspection of Fig. 4 shows that even for the most flat SPHC (C = -1 and +1), the PH decays towards both bounds of the envelope period.

The basic idea underlying our model approach is that SPHCs with different Cs result in PHs with different widths which in turn result in different ITD sensitivity. Note that our ROC analysis implicitly assumes that the process of ITD detection of our ideal observer is equivalent to the process of identifying which of the left and right ear signals is leading in time (the task the listeners actually performed in the experiments to be predicted). We assume that this simplification does not significantly affect the model's prediction power under the conditions of our study.

Our ROC model differs from the ROC analysis of binaural responses proposed by Shackleton et al. (2003). Their ROC analysis estimated ITD thresholds based on the neural firing rates measured from inferior colliculus neurons as a function of stimulus ITD. In contrast, our ROC analysis does not require specific properties of a binaural comparison unit because it evaluates the relative timing between monaural AN representations. Our approach is rather similar to the computations used to obtain "neurometric thresholds" in modulation detection (Johnson et al., 2012; Sayles et al., 2013).

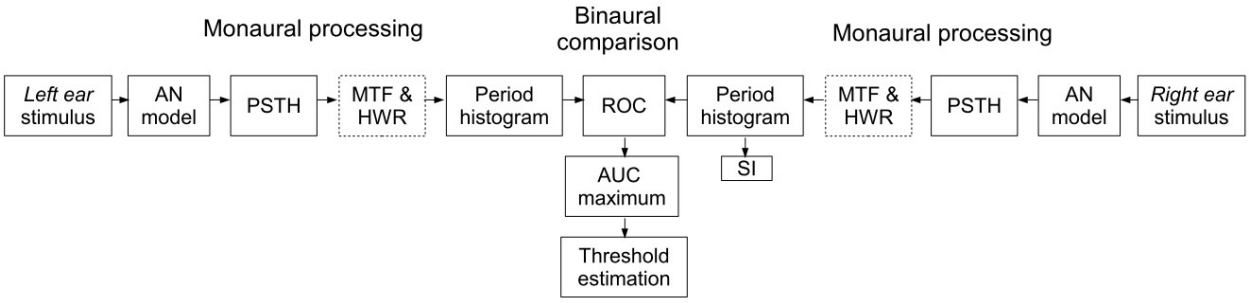

Fig. 7: Processing stages of the model used to predict ITD thresholds The monaural processing stages for the left and right ear feed the binaural comparison stage at the center of the figure. The estimation of ITD threshold is based on the maximum area under the ROC curve (AUC). The stages called modulation transfer function (MTF) and half-wave rectification (HWR) were added in the "revised" model version (see Figs. 13 to Fig. 15). See text for details of the model.

1. Model details

Figure 7 depicts the structure of our complete model, combining CF-matched monaural AN pre-processing for the left and right ear channels with a binaural comparison stage: the PSTHs from the left and right ear stimuli, optionally processed by a stage of modulation transfer function (MTF; see Sec. IV.C.5), are concatenated into PHs, which are then compared by means of the ROC analysis to finally estimate ITD thresholds. Fig. 8a shows two PHs with a relative delay (corresponding to the stimulus ITD) of 800 μs for SPHCs with C = -0.25, which are referred to as leading and lagging PHs, respectively. Fig. 8e shows example ROC curves constructed based on the cumulative probabilities of the lagging and leading PHs, from which the area under the ROC curve (AUC) is computed. For a fixed ITD (800 μs in this example), the binaural ROC model predicts lower AUC, i.e., lower ITD sensitivity, for flat stimulus envelopes (e.g., C = +1; thin solid line) compared to peaked stimulus envelopes (e.g., C = -0.25; thickest solid line), because the PHs for flat stimulus envelopes are flatter and therefore overlap more for a given ITD.

Before actually performing the ROC model analysis, the PHs have to be circularly shifted so that they are centered within the envelope period (10 ms for our stimuli) and thus

fulfill the requirement imposed by the ROC analysis. Fig. 8a shows the PHs already after the shifting operation. For comparison, Fig. 8c shows the unshifted versions of the PHs from Fig. 8a. Note that the amount of shift required to "center" the PHs depends on the stimulus phase curvature in combination with the monaural processing delay of the auditory periphery at the given CF. Figure 8d demonstrates how the AUC depends on the amount of shift (within the period) for this particular example stimulus. The AUC is apparently low when the histograms are not shifted (indicated by the diamond symbol), whereas the AUC is maximal for shifts approximately between 25 and 75 % of the period (the maximum indicated by the square symbol). While the shift required could be determined by some monaural criterion ensuring that the distributions are centered within the histogram window, we decided to determine it by maximizing the AUC. For each stimulus condition (i.e., each value of *C*), the optimal shift maximizing the AUC was determined based on the largest ITD considered (800 μs). This optimal AUC was then applied also for all other ITDs for that condition. For simplicity, by using the term AUC in the following, we refer to the maximum (optimized) AUC. It should be kept in mind that our ROC model is intended to represent an abstract optimum-observer binaural comparison stage, which, in reality, is most likely realized by coincidence detection neurons which require no shifting operation (temporal alignment). Thus, by determining the PH shift via maximizing the AUC, we attempt to avoid an arbitrarily unfavorable temporal alignment of the PH, which is a matter of computational modeling. Figure 8e shows the example condition of C = -0.25 and 800-μs ITD; optimizing the shift (according to the square symbol in Fig. 8d) gives the largest AUC (thickest solid line), whereas without the shift (according to the diamond symbol in Fig. 8d) the ROC curve has a "dent" that falls below the diagonal (thick dotted line). Considering only conditions including shift, decreasing the ITD from 800 to 200 μs reduces the AUC from 0.72 to 0.56 (the latter depicted with the medium-thick solid line in Fig. 8e). Finally, with zero ITD difference (thin dotted line), the ROC curve falls on the diagonal indicating chance performance

497 (AUC 0.5).

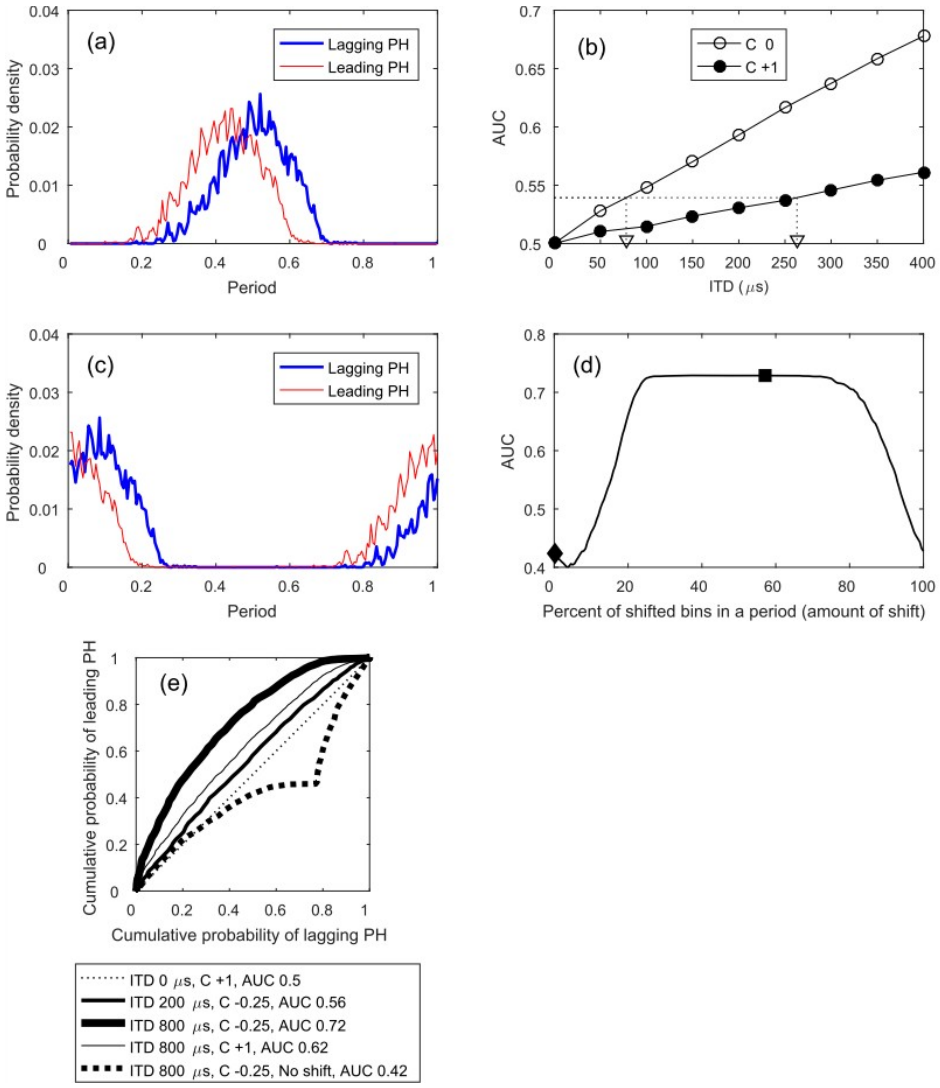

Fig. 8: (Color online) Illustration of the main steps of the binaural receiver operating characteristic (ROC) analysis. (a): Leading and lagging period histogram (PH) for C = -0.25, corresponding to the probability density functions of lagging AN responses (thick) with an ITD (800 μs in this example) and leading AN responses (thin). The x-axis is restricted to one period of the signal (i.e., 10 ms). The example PHs are shifted in order to maximize the area under the ROC curve (AUC). (b): AUC as a function of ITD for C = 0 and +1. The points were linearly interpolated and the inverse of the criterion AUC was defined as the predicted threshold (vertical arrows). The criterion AUC was found by minimizing the root-mean-square error (RMSE) between mean experimental thresholds and predicted thresholds. (c): The two histograms before applying the shifting operation, see text for details. (d): AUC as a function of the amount of shift, with the diamond showing the AUC without shift and the square showing the AUC with the optimal shift to maximize the AUC as in (a). (e): Examples of ROC curves with different ITDs and Cs with and without the optimal shift, see text.

499 Figure 8b shows the AUC as a function of ITD for example *C*s of 0 and +1. For visual

clarity, the functions are plotted only up to 400-μs ITD. For the threshold predictions, the AUC-vs-ITD functions were obtained for each of the nine *C*s and ITDs from 0 to 800 μs in 50-μs steps. The functions were then linearly interpolated. The arrows indicate predicted thresholds for the two *C*s at the given criterion AUC. In order to predict thresholds for the entire set of *C*s, the criterion AUC was systematically varied as the only free parameter to minimize the root-mean-square error (RMSE) between the mean experimental thresholds and corresponding predictions across all *C*s. All the RMSEs and criterion AUCs estimated in the present study are listed in Table I and Table II, respectively. Throughout this paper, RMSEs were calculated in the base-10 logarithmic ITD scale and the resulting minimum RMSEs were converted back to the linear ITD scale as similarly done by Klein-Hennig et al. (2011). Thresholds were predicted separately for the three fiber types and two OHC scaling factors. For completeness, Table I lists also the RMSEs for the predictions that were obtained by either linearly averaging across the three SR fiber types or by weighted averaging according to the prevalence of the three fiber types in cats (Liberman, 1978). Because we observed no consistent advantage of any type of averaging, in the following we report the predictions for the individual fiber types.

2. Predicted ITD thresholds based on normal OHCs

Figure 9 shows the predicted ITD thresholds as a function of *C* for different SR fibers, as compared with the mean thresholds from Experiment 1. The model clearly does not account for the threshold peak at $C = -0.5$, as already suggested by our monaural model analysis. We will reconsider this point in the next section. Besides the threshold peak, the model well predicts the overall pattern of thresholds. Most importantly, it correctly predicts the minimum threshold at $C = 0$ in the mean experimental data across listeners. The first and second rows of Table I indicate the RMSEs between the mean thresholds and predictions including and excluding $C = -0.5$ in the predictions, respectively. As expected, excluding the data point at $C = -0.5$ reduces the RMSEs, particularly for the low- and medium-SR fibers.

All three fiber types were found to predict the ITD thresholds about equally well. It was somewhat surprising that the high-SR fibers obviously conveyed sufficient temporal information to extract ITD cues, although the monaural SIs were found to be lower for the high-SR fibers compared to the low- and medium-SR fibers because of saturation (see Fig. 5a). Note, however, that saturation of high-SR fibers is shown to play a role when adding an additional stage of modulation filtering in the model (see Section IV.C.5 below).

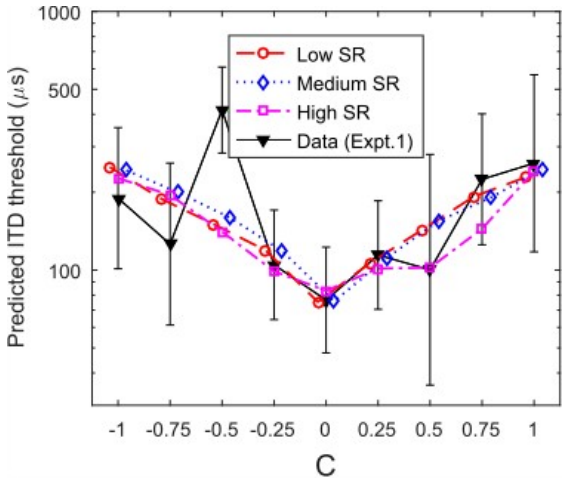

Fig. 9: (Color online) ITD threshold predictions as a function of C. The empty symbols connected by different dashed and dotted lines indicate the predictions of the initial binaural ROC model (without the MTF and HWR stage) for different fiber types. The filled triangles connected by the solid line denote the mean experimental thresholds replicated from Fig. 2. For clarity, some points are slightly shifted along the horizontal axis.

We considered that one possible explanation for the threshold peak could be an effect of listeners using AFs remote from the center frequency of the stimulus in addition to or instead of the on-frequency AFs. The model was therefore rerun at CFs in 1/3 octave steps from 2 to 8 kHz. The resulting predictions (not shown) revealed generally worse prediction accuracy for off-frequency CFs, with the 4-kHz CF showing the lowest RMSE. Most importantly, for none of the CFs the threshold peak at C = -0.5 was predicted by the model. Thus, off-frequency listening is

544 unlikely to contribute to or explain the non-monotonic threshold peak.

545 3. Predicted ITD thresholds of recorded stimuli

546 In search of an explanation for the unexpected ITD threshold peak at C = -0.5, we speculated that
547 the headphones used in our experiment might have produced some phase distortion that led to a
548 degradation of the envelope ITD cue for that particular condition. To test this idea by means of
549 modeling, monaural stimulus waveforms were recorded from the left and right headphones[2] and
550 ITDs were subsequently imposed to generate binaural stimuli. These binaural stimuli were used
551 as input to the binaural ROC model to predict ITD thresholds.

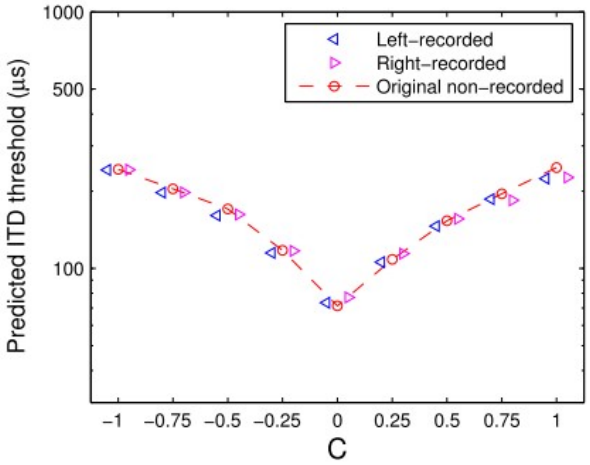

Fig. 10: (Color online) Predicted ITD thresholds from low-SR fibers as a function of C with headphone-recorded stimuli (left- and right-pointing triangles for left and right headphone outputs, respectively) and with original non-recorded stimuli (circles connected by dashed line). For clarity, the predictions of recorded stimuli are slightly shifted along the horizontal axis.

553 Figure 10 shows the predictions for low-SR fibers based on either the left- or the right
554 headphone recordings (left- and right-pointing triangles, respectively). They are almost the same
555 as those based on the original (non-recorded) stimuli (circles), without any indication of a non-
556 monotonic peak at C = -0.5. The same result was found for the medium and high-SR fibers (not
557 shown). Moreover, we observed that the crest factors of headphone recorded waveforms are

generally preserved compared to the digitally generated stimuli (not shown). Thus, the threshold peak is unlikely originating from the particular properties of the headphones.

4. Predicted ITD thresholds as a function of attack and pause duration

We tested the ROC model on literature data in order to evaluate its predictive ability in systematic manipulations of the temporal envelope shape. Klein-Hennig et al. (2011) systematically varied different aspects of the temporal envelope shape imposed on 4-kHz pure tones with an overall duration of 500 ms. The attack and decay flanks in each period were shaped by squared-sine functions, and the hold and pause segments had constant amplitudes at the desired level and at zero, respectively. We focus here on modeling the effects of independently varying the attack duration and the pause duration, as the importance of these envelope parameters was also suggested by the results of Laback et al. (2011). The same procedure as described in Sec. B.1 was used.

In their experiment on the effect of the attack duration, the hold and decay durations were kept constant, resulting in modulation rates between 35 and 50 Hz. Further stimulus details can be found in Table I of Klein-Hennig et al. (2011). Figure 11 shows their mean experimental thresholds across listeners and our model prediction as a function of attack duration. Both data and predictions show increasing thresholds as a function of attack duration, with our predictions falling within one standard deviation of the measured thresholds. In order to directly compare the performance of our binaural ROC model to Klein-Hennig et al.'s models, the RMSEs for their NCC model and for their NCC model with five adaptation loops (NCC5A) and one adaptation loop (NCC1A) are shown in the seventh to ninth rows of Table I. The RMSEs for our model are overall smaller compared to their models, particularly for the low- and medium-SR fibers. Overall, our model appears to well predict ITD thresholds across attack durations.

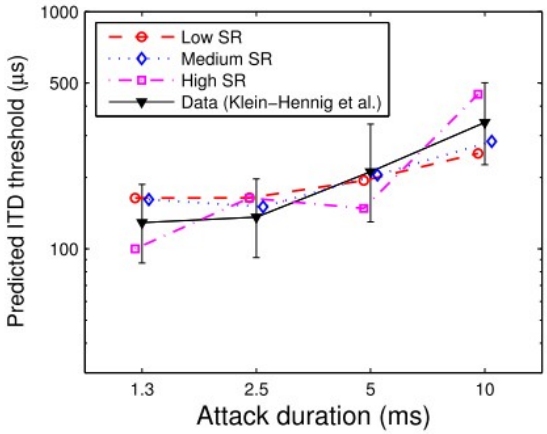

Fig. 11: (Color online) ITD thresholds and predictions for independent variation of the attack duration in each envelope cycle of amplitude-modulated 4-kHz tones. The filled triangles connected by the solid line denote the mean experimental thresholds (±1 standard deviation) reported in Klein-Hennig et al. (2011). The empty symbols indicate predictions of the original binaural ROC model (without the MTF and HWR stage) for different fiber types. For clarity, some points of prediction are slightly shifted along the horizontal axis.

In Klein-Hennig et al.'s experiment on the effect of pause duration, the attack duration, decay duration, and modulation rate were all kept constant, resulting in covarying hold duration (for details see Table I of Klein-Hennig et al.). Figure 12 shows the experimental thresholds together with our predictions as a function of pause duration for different fiber types[3]. In line with the experimental thresholds, the predicted thresholds decrease with increasing pause duration up to 8.8 ms. For longer pause durations, however, the predicted thresholds continue to decrease up to the longest pause duration tested (17.5 ms), while the experimental thresholds are almost constant. The models tested in Klein-Hennig et al. (2011) showed a similar tendency. We considered that these underestimations of predicted thresholds may be due to the lack of modulating filtering in the model. This idea is further motivated and evaluated in the next section.

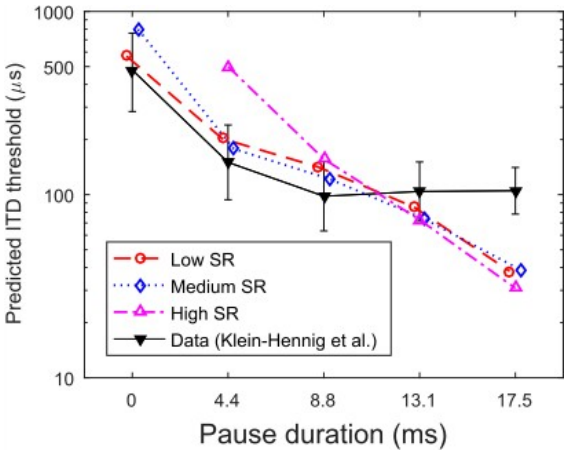

Fig. 12: (Color online) ITD thresholds and predictions for independent variation of the pause duration in each envelope cycle of amplitude-modulated 4-kHz tones. For other details, see caption for Fig. 11.

5. ITD thresholds predicted by the revised model

There are several indications from the physiological and psychophysical literature for some type of modulation filtering occurring prior to or at the level of binaural interaction (e.g., Yin and Chan 1990; Bernstein and Trahiotis, 2002; Wang et al., 2014). We supposed that such a modulation filtering may be required to obtain more similar neural envelope representations for pause durations beyond about 9 ms and, thus, better prediction of ITD thresholds for these stimuli from Klein-Hennig et al. (2011) (see Sec. IV.C.4). To that end, we added a modulation filter stage guided by the MTF measured in the medial superior olive (e.g., Yin and Chan 1990). Based on Wang et al. (2014), we combined a first-order low-pass and a third-order high-pass filter with 300-Hz cutoff frequencies each, together forming a band-pass filter. This filter stage was applied to the PSTHs (in units of spikes/sec). To avoid negative values in the PSTH, half-wave rectification (HWR) was added subsequent to the band-pass filtering (Joris, 1996; Nelson and Carney, 2004). Note that these two new stages are performed before calculating the PH. All the model predictions presented so far were reconsidered with the revised model (see Fig. 7 for

609 the model structure including the additional stage framed with a dotted line).

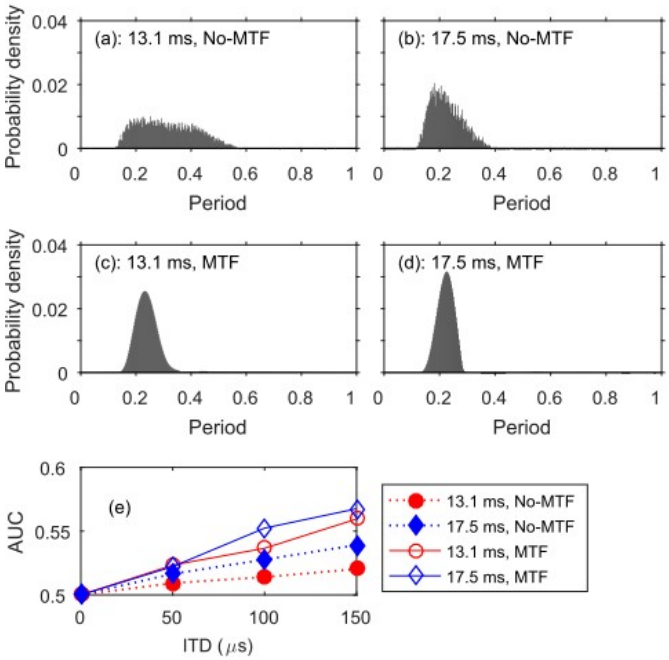

Fig. 13: (Color online) Period histograms (PHs) of low-SR fibers for stimuli with long pause durations, 13.1 and 17.5 ms, in each envelope cycle of amplitude-modulated 4-kHz tones as used Klein-Hennig et al. (2011). (a) and (b): PHs without MTF and half-wave rectification (HWR). (c) and (d): PHs with MTF and HWR. (e): AUC as a function of ITD for the pause durations 13.1 and 17.5 ms. The empty and filled symbols denote the AUCs with and without MTF and HWR, respectively.

611 We first evaluated the revised model on the pause-duration data. Figures 13a to 13d
612 show the PHs of low-SR fibers for the relevant long pause durations (13.1 and 17.5 ms) with and
613 without modulation filtering (and HWR). Without modulation filtering, the histogram for 17.5
614 ms (Fig. 13b) is more compact than the histogram for 13.1 ms (Fig. 13a), which explains the
615 lower predicted threshold for the 17.1-ms pause duration in Fig. 12. In contrast, when including
616 modulation filtering (Fig. 13c and 13d), the PHs are much more similar and generally more
617 narrow than those without modulation filtering. Figure 13e shows the AUC-vs-ITD functions
618 corresponding to the PHs shown in Fig. 13a to 13d. For clarity, the range of ITD is restricted
619 within 150 µs within which the predicted thresholds were anticipated. Consistent with the PHs,

625 the narrower PHs for the model with the modulation filter result in overall larger AUC values
626 and steeper AUC-vs-ITD functions. More importantly, the AUC-vs-ITD functions are similar for
627 the two pause durations with modulation filtering (empty symbols), whereas they differ more
628 without modulation filtering (filled symbols). As expected from these AUC-vs-ITD functions,
629 the predictions of the revised model (Fig. 14a) better represent the saturation of thresholds at
630 longer pause durations as compared to the original model (Fig. 12). Across all the pause
631 durations, the predictions fall almost within one standard deviation of the actual thresholds
632 across listeners. The revised model also produced valid threshold predictions for high-SR fibers.
633 The RMSEs are overall lower than those for the different model versions presented in Klein-
634 Hennig et al. (see Table I).

635 Fig. 14b (middle panel) shows the performance of the revised model for Klein-
636 Henning's data on the effect of attack duration. Compared to the original model version (Fig. 11),
637 the predictions are comparably accurate (see also RMSEs in Table I). This suggests that the
638 additional stage of band-pass filtering (and HWR) is not important for predicting the thresholds
639 across attack durations.

640 Figure 14c (bottom panel) shows the predictions of the revised model for our
641 thresholds across C obtained in Experiment 1. In case of low- and medium-SR fibers, the
642 predictions remain generally accurate, showing the minimum at $C = 0$. The predictions for high-
643 SR fibers show a pattern completely inconsistent with the experimental data. As shown in the
644 monaural SI calculations, high-SR fibers appear to be saturated for the particular stimulus level.
645 While this did not sufficiently impair the temporal code in the model version without the MTF,
646 the application of modulation filtering for the high SR fibers *increased* the relative peakedness of
647 the PH for $C = +1$ and -1, whereas it *decreased* the peakedness of the PH for $C = 0$ (not shown),
648 suggesting that the modulation filtering adversely degraded the temporal code of high SR fibers.
649 Note that the experiments by Klein-Hennig et al. (2011) were performed at lower stimulus levels

650 (between 60 and 65 dB SPL), which may explain the relatively better predictions for this fiber
651 type. Notably, as for the original model version, none of the fiber types in the revised model
652 accounted for the threshold peak at C = -0.5.

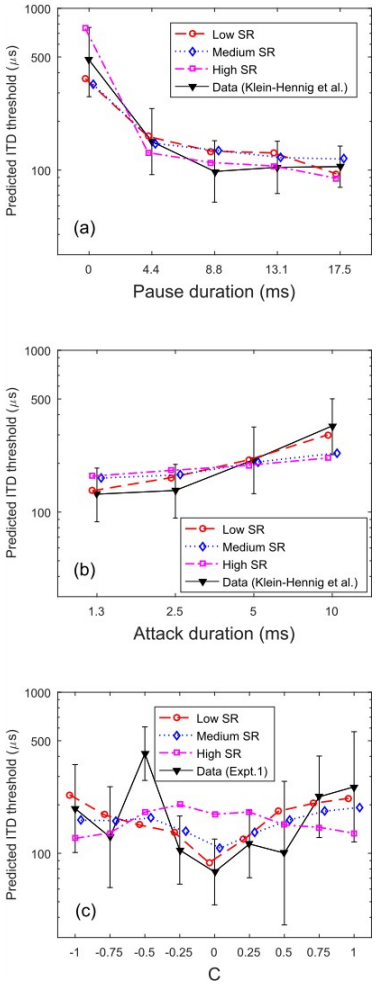

Fig. 14: (Color online) ITD threshold predictions of the "revised" binaural ROC model, which includes the MTF and HWR stage. (a) ITD thresholds as a function of pause duration (data from Klein-Hennig et al.). (b) ITD thresholds as a function of attack duration (data from Klein-Hennig et al.). (c) ITD thresholds as a function of C. For comparison, the corresponding predictions without the MTF and HWR stage are shown in Figs. 12, 11, and 9, respectively.

In summary, including a modulation filter allows to better account for a larger body of envelope ITD data, particularly when considering low-SR fibers. For high-SR fibers, the effect of the modulation filter on the prediction power is either beneficial (pause duration), marginal (attack duration), or largely harmful (current Experiment 1). The different stimulus levels in the different experiments appears to contribute to the differential effects of the modulation filter. While the temporal code in high-SR fibers appears to be degraded compared to low- and mid-SR fibers (as shown by the monaural SIs provided in Section IV.B.1), it appears to be just sufficient for extracting ITD cues without modulation filter (Figure 9). In contrast, the addition of the modulation filter seems to render those ITD cue unreliable (Figure 14c).

6. ITD threshold predictions with OHC loss using the revised model

In the Introduction, we hypothesized that the reduced compression associated with OHCs loss may enhance stimulus phase effects on ITD sensitivity. The monaural model analysis supported this idea by showing larger differences in SI across *C*s with OHC loss compared to normal OHCs (Fig. 5). We therefore speculated that predicted ITD thresholds would also differ by a greater amount across *C*s with OHC loss compared to normal OHCs. Because experimental data with HI listeners were not available to optimize the AUC criterion at threshold, we used the same AUC criterion, i.e., decision criterion, for each fiber type that was found to be optimal to predict the NH listeners' data (see also Table II). This assumption is not unreasonable given that binaural processing is often assumed to be normal in listeners with cochlear hearing loss.

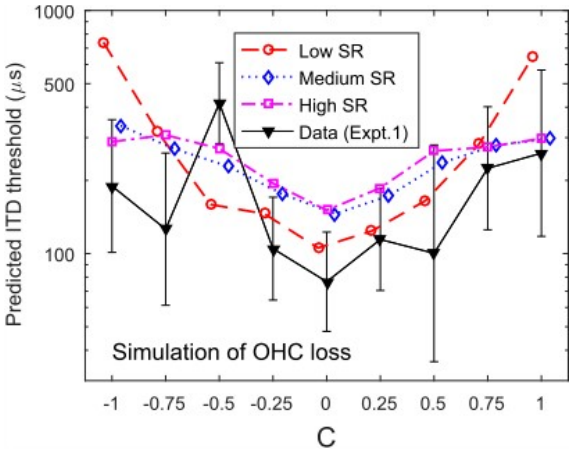

Fig. 15: (Color online) ITD threshold predictions as a function of C simulating complete OHC loss, using the "revised" binaural ROC model. The AUC criteria for the individual fiber types were taken from the corresponding simulations with normal OHCs in Fig. 14c (see also Table II).

Figure 15 shows the resulting predictions of ITD thresholds across *C* with OHC loss, using the revised model. Consistent with the SI predictions, for low-SR fibers, OHC loss results in a larger range of ITD thresholds across *C* as compared to normal OHCs (see corresponding predictions of the revised model in Fig. 14c). Medium- and high-SR fibers provide very similar patterns of predicted thresholds, with a markedly smaller range of thresholds compared to the low-SR fibers. Note that with normal OHCs the predicted threshold pattern for high-SR fibers was found to be completely inconsistent with the data. The relatively better predictability for high-SR fibers with OHC loss compared to normal OHCs supports the above-mentioned interpretation that OHC loss shifts the response level below the saturation region for the fiber's operation range at the given stimulus level. The thresholds predicted for OHC loss tend to be higher compared to those predicted for normal OHCs. This appears consistent with the overall higher SI values in normal OHCs, at least for the low- and medium-SR fibers, and with the overall higher spike rates. Thus, the predicted thresholds with OHC loss being overall larger than those with normal OHCs is broadly consistent with experimentally measured envelope ITD

thresholds for AM tones being larger in HI than in NH listeners (Lacher-Fougère and Demany, 2005). Overall, these simulations support the idea that the lack of compression in cochlear hearing loss may enhance the differences in ITD thresholds across *C*s, at least in low-SR fibers.

V. GENERAL DISCUSSION AND CONCLUSIONS

The goal of this study was to evaluate the idea that ITD thresholds vary as a function of the signal's phase curvature and thus may provide a measure of the phase response at the AF centered at the stimulus. To that end, Experiment 1 measured ITD thresholds using SPHCs centered at 4 kHz with various phase curvatures quantified by the parameter *C*. The frequencies of all spectral components were sufficiently high to ensure that they were not resolved and could not provide fine-structure ITD cues. The results showed significant and systematic effects of *C* on ITD thresholds across our test group of seven NH listeners. Given that only the phase properties of the harmonic complexes were manipulated, keeping the power spectrum constant, the effect can only be due to the temporal interaction of the stimulus' spectral components, resulting in envelope ITD cues. According to our hypothesis, the different amounts of envelope peakedness for the stimuli with different *C*s (Fig. 1) would result in different envelope ITD thresholds. As such, the reported ITD thresholds reflect the exclusive effect of changes in the temporal envelope shape, without any confounding changes of the long-time stimulus spectrum, which has, to our knowledge, not been shown before.

Our results potentially provide insight into the phase response of the AF centered at the stimulus. It was assumed that the "internal" envelope representation, i.e., after passing the AF, is important for the ITD extraction process and that ITD thresholds would be lowest for stimuli with the most peaked "internal" envelope. On average across listeners, the minimum ITD threshold was found at $C = 0$. According to the idea that the stimulus phase curvature resulting in

716 minimum ITD threshold reflects the mirrored phase curvature of the auditory periphery, this
717 would suggest that there is no phase-curvature conversion in the peripheral auditory system at
718 the CF centered at our stimuli (4 kHz), and the phase curvature is approximately zero. On a more
719 detailed level, the ITD data show a relatively large amount of uncertainty about the position of
720 the minimum along the *C* scale both within and across listeners, which could be somewhere
721 between -0.25 and 0.5. In contrast, masking data from an experiment involving similar stimuli
722 and largely the same listeners (Tabuchi et al., 2016) revealed a more robust minimum that
723 occurred between $C = 0.25$ and 0.5. Together, these results seem to suggest that the current ITD
724 paradigm does not provide sufficiently stable estimates of the minimum in ITD sensitivity across
725 *C*s that could be used to infer the cochlear phase response in individual listeners.
726 An unexpected non-monotonous threshold elevation (peak) was consistently observed
727 for a particular phase curvature, namely at $C = -0.5$. Because this peak was rather unexpected
728 (we are not aware of any masking data using SPHC stimuli showing such a peak), we took some
729 effort to evaluate and understand its origin. First, in a follow-up experiment with two of the
730 listeners using a finer sampling of *C*s (Experiment 2), we replicated the presence of the peak, and
731 found even a secondary peak in one listener. Second, we ruled out a potentially confounding
732 influence of the headphones' transfer function. Third, we performed a modeling analysis[4] to
733 evaluate if the combination of the auditory periphery model up to the level of the AN (Zilany et
734 al, 2014) and a probabilistic interaural comparison stage can predict the pattern of experimental
735 data. It was shown that model predicts the overall pattern of results, i.e., an essentially V-shaped
736 pattern, but it does not predict the peak at $C = -0.5$. As an additional check, we inspected the
737 temporal response to SPHCs with various *C*s of a time-domain cochlear model which has been
738 shown to well predict human otoacoustic emission data (Verhulst and Dau, 2012). The temporal
739 envelope peakedness of the model responses showed a very similar dependency on *C* as we
740 observed for the responses of the AN model used as a front-end of our ITD model, with no

741 indication for a non-monotonic behavior at *C*s around -0.5. In summary, assuming no errors in
742 our experimental setup (which we checked carefully), and assuming the general appropriateness
743 of the model (see below), all these results appear to suggest that the AF phase response at 4 kHz
744 may have some yet undiscovered feature producing a blurry "internal" envelope for *C*s around
745 -0.5. While the phase response of the AN front-end model has been fitted to AN data from the cat
746 (Zilany and Bruce, 2006), it is conceivable that the human phase response differs in some aspects
747 which affects the salience of the "internal" envelope ITD cues but not the amount of masking for
748 certain negative stimulus phase curvatures. Note that direct and non-invasive measurement of the
749 cochlear phase response in humans is difficult (see e.g., Paredes Gallardo et al., 2016).
750 In order to evaluate the wider applicability of our model, we predicted published data
751 on the effect of systematic variation of basic envelope shape parameters on envelope ITD
752 sensitivity (Klein-Hennig et al., 2011). Data on the effect of the attack duration within each
753 modulation cycle were well predicted by the model, whereas data on the effect of the pause
754 duration within the modulation cycle were only partly predicted. For large pause durations (> 9
755 ms), the model systematically underestimated the measured thresholds. Considering that this
756 discrepancy may be due to an improper envelope representation available to the interaural
757 comparison stage of our model, we included an envelope modulation filter stage which was
758 inspired by band-pass-like "MTF" characteristics of binaural envelope processing (Yin and
759 Chan, 1990; Wang et al., 2014) and follows also the general idea of other established auditory
760 models of some type of envelope filtering (e.g., Ewert and Dau, 2000). Adding this stage clearly
761 improved the predictability of the pause-duration data and predicted the data of attack-duration
762 and *C*s as well as the model without the modulation filter. The revised model provided slightly
763 better predictions than the well-established NCC model as well as a variant of it including a stage
764 simulating peripheral adaptation (Klein-Hennig et al., 2011). Interestingly, while the NCC model
765 has been shown to be most successful in predicting envelope ITD data when including a 150-Hz

766 modulation low-pass filter (e.g., Bernstein and Trahiotis, 2002), including such an envelope filter
767 in our model actually reduced the predictability of the pause-duration data. The more
768 physiology-based MTF stage included in our revised model has a clearly higher cutoff-frequency
769 of the low-pass filter component (300 Hz). The properties of the modulation filter should be
770 further evaluated and optimized based on a larger body of experimental data.

771 The ITD model in its current form is applicable only for periodic signals because it is
772 based on a probabilistic analysis of the PH. For application with non-periodic stimuli, the PH
773 stage could be replaced, for example by a shuffled cross-correlogram analysis (Joris, 2003). A
774 recent ongoing study (Prokopiou et al., 2015) using such an approach reported successful
775 prediction of the effects of envelope-shape manipulations reported in Laback et al. (2011),
776 similar to those tested in Klein-Hennig et al. (2011).

777 One motivation for the present study was to evaluate the general feasibility of a method
778 to measure the AF phase response, applicable also in listeners suffering from cochlear hearing
779 impairment characterized by OHC loss (Oxenham and Bacon, 2003). While the established
780 masking paradigm for determining the AF phase response relies on fast-acting cochlear
781 compression (e.g., Oxenham and Dau, 2004; Tabuchi et al., 2016) which is absent or reduced in
782 case of OHC loss, the ITD-based approach investigated in the present study does not rely on that
783 requirement. Our model predictions showed that OHC loss actually increases the effect of the
784 stimulus phase curvature on ITD thresholds, as compared to NH listeners. Our monaural SI
785 analysis (Fig. 5) suggested that this can be attributed to two effects: for low- and medium-SR
786 fibers, OHC loss decreases the SI for *C*s approaching -1 or +1 and has no effect for *C*s close to
787 zero, together enhancing the effect across *C*s. For high-SR fibers, in contrast, OHC loss enhances
788 the SI for *C*s close to zero and has no effect for *C*s approaching -1 or +1. Based on these
789 predictions, ITD thresholds may be expected to vary more across *C* in actual HI listeners
790 compared to NH listeners. Thus, using some type of ITD-based method may be a viable

approach to determine the phase response of AFs in cochlear hearing loss.

An interesting observation from our modeling using the "revised" version of the ITD model was that the temporal code provided by high-SR fibers for our medium-level (70 dB SPL) stimuli appears to be susceptible to response saturation effects, resulting in unreliable ITD threshold predictions for that fiber type. In contrast, the low- and medium-SR fibers provided much more accurate ITD threshold predictions, obviously because they were not affected by response saturation effects at the given stimulus level. Listeners may thus focus on low- and medium-SR fibers to extract ITD cues from these stimuli. It may, however, pose a problem for listeners suffering from noise-induced hearing loss where low- and mid-SR fibers seem to be predominantly impaired (Furman et al., 2013). Future experiments with listeners having cochlear hearing loss with different origins might address these issues.

Before extending the approach of the present study to actual HI listeners in future investigations, the open issues regarding the ITD paradigm mentioned above need to be addressed. A reliable measure of the AF phase response may make it then possible to incorporate the subject-specific phase response in the signal transmission of hearing devices in order to enhance the salience of temporal cues, most importantly pitch and ITD. Enhanced access to pitch and ITD cues would likely improve auditory communication in challenging listening environments involving multiple sound sources.

VI. ACKNOWLEDGMENTS

We thank two anonymous reviewers for providing helpful comments on an earlier version of the manuscript. We also thank Michael Mihocic for his assistance in writing software programs for the experiments (*ExpSuite*), Katharina Zenke for her assistance in collecting preliminary data, and Dr. Piotr Majdak and Dr. Thibaud Necciari for helpful comments. This work was supported by the Austrian Science Fund (FWF, P24183-N24).

FOOTNOTES

1. Matlab code of a Markov chain Monte Carlo technique (Fründ et al., 2011) is available at http://psignifit.sourceforge.net/.

2. For the recordings, the headphones stimuli were sent through an artificial ear (4153, Bruel & Kjær) and a sound level meter (2260, Bruel & Kjær), and digitized by a sound interface (E-Mu 0404, Creative Professional) at a sampling rate of 48 kHz. Recordings were made using the free audio software Audacity (version 2.11; http://audacityteam.org/).

3. For the high-SR fibers, at the pause duration of 0 ms the AUC-vs-ITD function was completely flat around AUC 0.5 and thus did not allow to estimate a threshold as shown in Fig. 12.

REFERENCES

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise," J Acoust Soc Am **95**, 3561-3567.

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. (**2009**). "How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli," J Acoust Soc Am **125**, 3234-3242.

Brughera, A., Dunai, L., and Hartmann, W. M. (**2013**). "Human interaural time difference thresholds for sine tones: the high-frequency limit," J Acoust Soc Am **133**, 2839-2855.

Carlyon, R. P., and Datta, A. J. (**1997**). "Excitation produced by Schroeder-phase complexes: evidence for fast-acting compression in the auditory system," J Acoust Soc Am **101**, 3636-3647.

Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J Acoust Soc Am **93**, 401-417.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). “A quantitative model of the ‘effective’ signal processing in the auditory system. I. Model structure,” J Acoust Soc Am **99**, 3615–3622.

Dietz, M., Klein-Hennig, M., and Hohmann, V. (**2015**). "The influence of pause, attack, and decay duration of the ongoing envelope on sound lateralization," J Acoust Soc Am **137**, EL137-143.

Dietz, M., Wang, L., Greenberg, D., and McAlpine, D. (**2016**). "Sensitivity to interaural time differences conveyed in the stimulus envelope: estimating inputs of binaural neurons through the temporal analysis of spike trains," J Assoc Res Otolaryngol **17**, 313-330.

Dreyer, A., and Delgutte, B. (**2006**). "Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: implications for sound localization," J Neurophysiol **96**, 2327-2341.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J Acoust Soc Am **108**, 1181-1196.

Francart, T., Lenssen, A., and Wouters, J. (**2012**). "The effect of interaural differences in envelope shape on the perceived location of sounds (L)," J Acoust Soc Am **132**, 611-614.

Fründ, I., Haenel, N. V., and Wichmann, F. A. (**2011**). "Inference for psychometric functions in the presence of nonstationary behavior," J Vis **11**, 1-19.

Furman, A. C., Kujawa, S. G., and Liberman, M. C. (**2013**). "Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates," J Neurophysiol **110**, 577-586.

Gai, Y., Kotak, V. C., Sanes, D. H., and Rinzel, J. (**2014**). "On the localization of complex sounds: temporal encoding based on input-slope coincidence detection of envelopes," J Neurophysiol **112**, 802-813.

Goldberg, J. M., Brown, P. B. (**1969**). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization," J. Neurophysiol **32**, 613-636.

Green, D. M., and Swets, J. A. (**1974**). *Signal detection theory and psychophysics* (Krieger, New York).

Grothe, B., Pecka, M., and McAlpine, D. (**2010**). "Mechanisms of sound localization in mammals," Physiol Rev **90**, 983-1012.

Hanley, J. A., and McNeil, B. J. (**1982**). "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology **143**, 29-36.

Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). "Auditory nerve model for predicting performance limits of normal and impaired listeners," ARLO 2, 91-96.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J Acoust Soc Am **55**, 84-90.

Henry, K. S., and Heinz, M. G. (**2013**). "Effects of sensorineural hearing loss on temporal coding of narrowband and broadband signals in the auditory periphery," Hear Res **303**, 39-47.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J Acoust Soc Am **68**, 1115-1122.

Johnson, J. S., Yin, P., O'Connor, K. N., and Sutter, M. L. (**2012**). "Ability of primary auditory cortical neurons to detect amplitude modulation with rate and temporal codes: neurometric analysis," J Neurophysiol **107**, 3325-3341.

Joris, P. X. (**1996**). "Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive," J Neurophysiol **76**, 2137-2156.

Joris, P. X. (**2003**). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," J Neurophysiol **23**, 6345-6350.

Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J Acoust Soc Am **91**, 215-232.

Kale, S., and Heinz, M. G. (**2010**). "Envelope coding in auditory nerve fibers following noise-induced hearing loss," J Assoc Res Otolaryngol **11**, 657-673.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J Acoust Soc Am **97**, 1817-1829.

Klein-Hennig, M., Dietz, M., Hohmann, V., and Ewert, S. D. (**2011**). "The influence of different segments of the ongoing envelope on sensitivity to interaural time delays," J Acoust Soc Am **129**, 3856-3872.

Laback, B., Zimmermann, I., Majdak, P., Baumgartner, W. D., and Pok, S. M. (**2011**). "Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing," J Acoust Soc Am **130**, 1515-1529.

Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J Acoust Soc Am **118**, 2519-2526.

Lentz, J. J., and Leek, M. R. (**2001**). "Psychophysical estimates of cochlear phase response: masking by harmonic complexes," J Assoc Res Otolaryngol **2**, 408-422.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J Acoust Soc Am **63**, 442-455.

Myung, I. J. (**2003**). "Tutorial on maximum likelihood estimation," J Math Psychol **47**, 90-100.

Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J Acoust Soc Am **116**, 2173-2186.

Oxenham, A. J., and Bacon, S. P. (**2003**). "Cochlear compression: perceptual measures and implications for normal and impaired hearing," Ear Hear **24**, 352-366.

Oxenham, A. J., and Dau, T. (**2001**). "Reconciling frequency selectivity and phase effects in masking," J Acoust Soc Am **110**, 1525-1538.

Oxenham, A. J., and Dau, T. (**2004**). "Masker phase effects in normal-hearing and hearing-impaired listeners: evidence for peripheral compression at low signal frequencies," J Acoust Soc Am **116**, 2248-2257.

Oxenham, A. J., and Ewert, S. D. (**2005**). "Estimates of auditory filter phase response at and below characteristic frequency (L)," J Acoust Soc Am **117**, 1713-1716.

Paredes Gallardo, A., Epp, B., and Dau, T. (**2016**). "Can place-specific cochlear dispersion be represented by auditory steady-state responses?," Hear Res **335**, 76-82.

Patterson, R. D., and Irino, T. (**1998**). "Modeling temporal asymmetry in the auditory system," J Acoust Soc Am **104**, 2967-2979.

Prokopiou, A. N., Wouters, J., Francart, T. (**2015**). "Functional modelling of neural interaural time difference coding for bimodal and bilateral cochlear implant simulation", Conference on Implantable Auditory Protheses (CIAP): Tahoe City, CA, USA.

Sayles, M., Füllgrabe, C., and Winter, I. M (**2013**). "Neurometric amplitude-modulation detection threshold in the guinea-pig ventral cochlear nucleus," J Physiol **591**, 3401-3419.

Schroeder, M. (**1970**). "Synthesis of low peak-factor signals and binary sequences of low autocorrelation," IEEE Trans Inf Theory **16**, 85-89.

Shackleton, T. M., Skottun, B. C., Arnott, R. H., and Palmer, A. R. (**2003**). "Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of Guinea pigs," J Neurosci **23**, 716-724.

Shera, C. A. (**2001**). "Frequency glides in click responses of the basilar membrane and auditory nerve: their scaling behavior and origin in traveling-wave dispersion," J Acoust Soc Am **109**, 2023-2034.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc Natl Acad Sci U S A **99**, 3318-3323.

Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). "Phase effects in masking related to dispersion in the inner ear," J Acoust Soc Am **80**, 1631-1637.

Summers, V. (**2000**). "Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes," J Acoust Soc Am **108**, 2307-2317.

Summers, V. (**2001**). "Overshoot effects using Schroeder-phase harmonic maskers in listeners with normal hearing and with hearing impairment," Hear Res **162**, 1-9.

Summers, V., and Leek, M. R. (**1998**). "Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners," Hear Res **118**, 139-150.

Tabuchi, H., Laback, B., Necciari, T., and Majdak, P. (**2016**). "The role of compression in the simultaneous masker phase effect," J Acoust Soc Am **140**, 2680-2694.

Verhulst, S., Dau, T, and Shera C. (**2012**). "Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission," J Acoust Soc Am **132**, 3842-3848.

Wang, L., Devore, S., Delgutte, B., and Colburn, H. S. (**2014**). "Dual sensitivity of inferior colliculus neurons to ITD in the envelopes of high-frequency sounds: experimental and modeling study," J Neurophysiol **111**, 164-181.

Wojtczak, M., and Oxenham, A. J. (**2009**). "On- and off-frequency forward masking by Schroeder-phase complexes," J Assoc Res Otolaryngol **10**, 595-607.

Yin, T. C., and Chan, J. C. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J Neurophysiol **64**, 465-488.

Yost, W. A. (**1976**). "Lateralization of repeated filtered transients," J Acoust Soc Am **60**, 178-181.

Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J Acoust Soc Am **66**, 1381-1403.

Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," J Acoust Soc Am **109**, 648-670.

Zilany, M. S., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J Acoust Soc Am **120**, 1446-1466.

Zilany, M. S., and Bruce, I. C. (**2007**). "Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats," J Acoust Soc Am **122**, 402-417.

Zilany, M. S., 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., 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.

Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J Acoust Soc Am **28**, 860-864.

TABLES

Table I. Root-mean-square error (RMSE) in μs between the mean thresholds and model predictions. MTF and No-MTF refer to the binaural ROC models with and without modulation filtering and half-wave rectification, respectively. The RMSEs for the prediction averaged over the three SR fiber types are listed in the column "Mean SR", whereas the RMSEs for the weighted average according to the prevalence of the three fiber types in cats (Low SR: 16%, Medium SR: 23%, High SR: 61%; Liberman, 1978) are listed in the column "Weighted mean SR". The upper four rows show the RMSEs for the prediction across *C*s from the present study, whereas the lower rows show the RMSEs for attack and pause duration experiments in Klein-Hennig et al. (2011). The RMSEs for the normalized cross-correlation coefficient (NCC) model, the NCC model with five adaptation loops (NCC5A) and with the first adaptation loop only (NCC1A) were replicated from Table I in Klein-Hennig et al. (2011). The parenthesized criterion indicates that the ITD prediction of high SR fibers for the 0-ms pause duration was infinitely large due to the completely flat AUC-vs-ITD function, which prevented the systematic estimate.

TABLE I.

Stimulus	ITD Model	Fig.	Low SR	Medium SR	High SR	Mean SR	Weighted mean SR	Klein-Hennig et al. (2011)
C	ROC, No-MTF	9	1.50	***1.49***	1.53	1.49	1.50	
C excluding -0.5	ROC, No-MTF		***1.22***	1.24	1.27	1.23	1.24	
C	ROC, MTF	14c	1.54	***1.50***	1.78	1.55	1.63	
C excluding -0.5	ROC, MTF		***1.22***	1.31	1.70	1.36	1.49	
Attack	ROC, No-MTF	11	1.24	***1.17***	1.32	1.13	1.17	
Attack	ROC, MTF	14b	***1.12***	1.29	1.35	1.24	1.29	
Attack	NCC							1.62
Attack	NCC1A							1.31
Attack	NCC5A							1.56
Pause	ROC, No-MTF	12	***1.74***	1.84	(2.48)	(1.91)	(2.11)	
Pause	ROC, MTF	14a	***1.23***	1.24	1.26	1.13	1.14	
Pause	NCC							1.34
Pause	NCC1A							1.37
Pause	NCC5A							1.43

Table II. The criterion AUCs best predicting the ITD thresholds. The AUCs for the stimuli used in the current study and Klein-Hennig et al. (2011) are shown above in the upper four and lower four rows, respectively. See also the caption of Table I.

TABLE II.

Stimulus	ITD Model	Fig.	Low SR	Medium SR	High SR
C	ROC, No-MTF	9	0.534	0.531	0.513
C excluding -0.5	ROC, No-MTF		0.530	0.527	0.513
C	ROC, MTF	14c	0.542	0.546	0.537
C excluding -0.5	ROC, MTF		0.534	0.542	0.533
Attack	ROC, No-MTF	11	0.514	0.516	0.519
Attack	ROC, MTF	14b	0.556	0.577	0.578
Pause	ROC, No-MTF	12	0.517	0.512	(0.521)
Pause	ROC, MTF	14a	0.549	0.563	0.553