A framework for computational modelling of interaural time difference discrimination of normal and hearing-impaired listeners

Moncada-Torres, Arturo ; Joshi, Suyash Narendra; Prokopiou, Andreas; Wouters, Jan; Epp, Bastian; Francart, Tom

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
[10.1121/1.5051322](https://doi.org/10.1121/1.5051322)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Moncada-Torres, A., Joshi, S. N., Prokopiou, A., Wouters, J., Epp, B., & Francart, T. (2018). A framework for computational modelling of interaural time difference discrimination of normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 144(2), 940–954. <https://doi.org/10.1121/1.5051322>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A framework for computational modelling of interaural time difference discrimination of normal and hearing-impaired listeners

Arturo Moncada-Torres,^{1,a)} Suyash N. Joshi,² Andreas Prokopiou,¹ Jan Wouters,¹ Bastian Epp,² and Tom Francart¹

¹*KU Leuven – University of Leuven, Department of Neurosciences, ExpORL, Herestraat 49, Bus 721, 3000 Leuven, Belgium*

²*Department of Electrical Engineering, Hearing Systems, Technical University of Denmark, Ørstedes Plads, Building 352, DK-2800 Kongens Lyngby, Denmark*

(Received 19 October 2017; revised 1 August 2018; accepted 3 August 2018; published online 22 August 2018)

Different computational models have been developed to study the interaural time difference (ITD) perception. However, only few have used a physiologically inspired architecture to study ITD discrimination. Furthermore, they do not include aspects of hearing impairment. In this work, a framework was developed to predict ITD thresholds in listeners with normal and impaired hearing. It combines the physiologically inspired model of the auditory periphery proposed by Zilany, Bruce, Nelson, and Carney [(2009). *J. Acoust. Soc. Am.* **126**(5), 2390–2412] as a front end with a coincidence detection stage and a neurometric decision device as a back end. It was validated by comparing its predictions against behavioral data for narrowband stimuli from literature. The framework is able to model ITD discrimination of normal-hearing and hearing-impaired listeners at a group level. Additionally, it was used to explore the effect of different proportions of outer- and inner-hair cell impairment on ITD discrimination. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5051322>

[MD]

Pages: 940–954

I. INTRODUCTION

The best known theory of how we process interaural time differences (ITDs) was proposed by Jeffress (1948). In short, it suggests that ITDs are decoded by neurons in the central auditory system, which receive inputs from both ears and are sensitive to specific delays of their inputs. Thus, they function as coincidence detectors and fire at a particular ITD.

Different computational frameworks based on the Jeffress model have been developed to understand, study, and simulate ITD perception. Initially, binaural phenomena were predicted using cross-correlation of the (continuous) signals arriving to the left and right ears (Sayers and Cherry, 1957). Later on, models started incorporating the correlation of the neural response of both ears at the level of the auditory nerve (AN) as a main component. For instance, Colburn and Latimer (1978) and Colburn (1973, 1977) used a coarse model of the AN (consisting of a bandpass filter, automatic gain control, and an exponential rectifier) together with a binaural display to predict ITD discrimination in tone bursts across different levels, as well as binaural unmasking. Stern and Colburn (1978), Stern and Trahiotis (1992), and Stern and Shear (1996) extended this model by including a central-ity weighting function (designed to model the larger proportion of coincidence detection units sensitive to shorter ITDs) and a straightness weighting function (designed to emphasize the internal delays consistent across different

frequencies). Bernstein and Trahiotis (2012), Lindemann (1986), and Stern and Colburn (1978) also combined ITD with interaural level differences (ILD) information to predict perceptual lateralization as well as binaural detection.

These frameworks have contributed greatly in the studying and understanding of the processing of ITDs by the auditory system. However, their approach for modelling the AN can be considered elemental, since they include a limited amount of aspects regarding its biology. Incorporating the anatomy and the physiological processes underlying the auditory system can provide a better insight into how it processes binaural sounds. More recently, different models have been developed that explicitly include some biological aspects of binaural auditory temporal processing. For instance, Gai *et al.* (2014) combined a model of the AN with a model of bushy cells to study ITD coding in medial superior olive (MSO) neurons. Takanen *et al.* (2014) developed a method for visualizing binaural interactions (including sound lateralization using ITDs) using models of the superior olivary complex. Wang *et al.* (2014) presented a model of the binaural pathway capable of simulating ITD sensitivity of an inferior colliculus (IC) neuron to sinusoidally amplitude modulated tones and broadband noise. However, there are few physiologically inspired models that have been used to investigate ITD discrimination (e.g., Brughera *et al.*, 2013; Hancock and Delgutte, 2004; Prokopiou *et al.*, 2017).

While substantial progress has been achieved in the study of how normal-hearing (NH) listeners process ITD cues (Bernstein, 2001; Blauert, 1997; Colburn *et al.*, 2006; Grothe *et al.*, 2010; McAlpine, 2005; Stecker and Gallun,

^{a)}Electronic mail: arturo.moncadatorres@med.kuleuven.be

2012; Wang and Brown, 2005), this is not the case for hearing-impaired (HI) listeners (Akeroyd and Whitmer, 2016; Durlach *et al.*, 1981; Moore, 2007). HI listeners frequently present large inter-subject variability (Gabriel *et al.*, 1992; Hawkins and Wightman, 1980; Smoski and Trahiotis, 1986; Spencer *et al.*, 2016). Furthermore, it is hard to discern if the HI listeners' binaural hearing ability was affected because of HI itself or because of confounding factors, such as effort, concentration on the task, or age (Gallun *et al.*, 2014; King *et al.*, 2014; Peelle and Wingfield, 2016).

Studying HI using a physiologically based computational approach (Durlach *et al.*, 1981; Moore, 1996) would allow us to systematically investigate the mechanisms that are detrimental to the (binaural) hearing system and the effect of specific aspects of HI in listeners' performance in different tasks without the confounds of behavioral experiments. Moreover, it could provide us with additional means to improve the design, implementation, and fitting of hearing devices (e.g., hearing aids, cochlear implants). Furthermore, there has been no attempt to model ITD discrimination in HI listeners using a detailed physiologically based approach.

In this work, we developed a computational framework that addresses these issues. Namely, we used our framework to study ITD discrimination by predicting its thresholds (i.e., just noticeable difference) in both NH and HI listeners. It combines the *physiologically inspired front end* proposed by Zilany *et al.* (2014, 2009) with a coincidence detection stage and a *neurometric-based decision device* as a *back end*. We validated it by comparing its ITD threshold predictions for different stimuli against behavioral data

reported in literature. Additionally, we used it to explore the effect of different proportions of outer- and inner-hair cell (OHC, IHC) impairment on ITD discrimination.

II. FRAMEWORK DESCRIPTION

The proposed framework is composed of three main building blocks: (1) auditory periphery (responsible for generating spike trains elicited by a given stimulus with an imposed ITD), (2) coincidence detection (its purpose is to compute the imposed ITD), and (3) decision device (its objective is to predict an ITD threshold based on the distributions of the computed ITDs). The framework's block diagram is shown in Fig. 1 and is explained in detail below.

A. Auditory periphery

As a front end, we chose the model proposed by Zilany *et al.* (2014, 2009). It is capable of reproducing response properties of AN fibers to acoustic stimuli and has been validated with physiological data over a wider range than previously existing models. Furthermore, it has been successfully used to study a variety of auditory phenomena, such as masking release (Bruce *et al.*, 2013), frequency selectivity (Jennings and Strickland, 2012), neural adaptation to sound level (Zilany and Carney, 2010), sensory responses to musical consonance-dissonance (Bidelman and Heinz, 2011), speech intelligibility in noise (Moncada-Torres *et al.*, 2017), neural coding of chimaeric speech (Heinz and Swaminathan, 2009), and overshoot adaptation (Jennings *et al.*, 2011).

It is composed of different modules, each simulating a particular element of the auditory periphery. For each

FIG. 1. (Color online) Block diagram of the proposed framework. Signals coming from the left and right ear (with an imposed ITD) are processed by the model of the auditory periphery proposed by Zilany *et al.* (2014) and Zilany *et al.* (2009). The resulting spike trains are fed to a coincidence detection stage, which computes the imposed ITD. This process is repeated numerous times in order to generate a distribution of the computations for different ITD values. These distributions are used as an input to the decision device, yielding as an output a predicted threshold value in μ s.

channel (left and right), the model received as an input an instantaneous pressure waveform with a sampling frequency of 100 kHz. In all cases, the delay was applied to the right waveform, shifting it in time with respect to the left waveform. For each ear, the stimulus was first passed through a filter emulating the middle ear. Then, the output was passed through a signal path and a control path. The signal path simulated the behavior of the OHC-controlled filtering properties of the basilar membrane in the cochlea and the transduction properties of the IHCs by a succession of non-linear and low-pass filters. The control path simulated the function of the OHCs in controlling basilar membrane filtering. It did so with a wideband basilar membrane filter followed by a non-linearity module and an OHC low-pass filter. The control path output fed back into itself and into the signal path, as well. The IHCs output then went through an IHC-AN synapse module with two power-law adaptation paths (which account for slow and fast adaptations) and a spike generator (which accounts for the activity, adaptation, and refractoriness of the AN response).

The model allows to be tuned with different parameters, which were defined as follows. We chose to simulate human AN fibers with the same characteristic frequency (CF) as the stimulus frequency (in the case of pure tones) or center frequency (in the case of bandpass noise). These were tuned using values reported by [Shera et al. \(2002\)](#). For the sake of

simplicity, we decided to simulate only high spontaneous rate (100 spikes/s) fibers, since in our previous work ([Prokopiou et al., 2017](#)) we found that the use of a physiologically relevant mixture of high-, middle-, and low-spontaneous rate fibers had no significant effects on temporal coding at the simulated levels (Table I). We opted for an implementation of the power-law synapse function given it yields a more physiologically accurate response to relatively long stimuli (although at the expense of higher computational power, [Zilany et al., 2009](#)). Last, we chose to use a pre-determined fixed seed for the fractional Gaussian noise to better separate the (independent) effects of internal (i.e., physiological) and external (i.e., stimulus-driven) noise ([Zilany et al., 2014](#)).

Another important parameter was the number of spike trains that needed to be simulated in order to obtain a physiologically relevant response for each ITD computation. We defined this number based on the work by [Louage et al. \(2006\)](#). They recorded spike trains from AN and anteroventral cochlear nucleus neurons of cats to investigate the temporal coding of sound to noise stimuli in afferent neurons and reported collecting ~ 3000 spikes per token. Therefore, we made sure that each simulated response had at least 3000 spikes for each channel (left and right). Considering the average number of generated spikes per second to each particular stimulus and the stimulus duration (Sec. III A), the

TABLE I. Participants and stimuli characteristics of the data sets used in the validation and exploration of the proposed framework. Conditions where parameters are shown in italics were simulated by the framework, but have no behavioral counterpart. Ref. = reference, av. = average, PT = pure tone, BPN = bandpass noise, thr. = threshold, dur. = duration, dec. dev. = decision device.

Participants					Stimuli					Dec. dev.
Ref.	Hearing status	<i>N</i>	Age [yr]	Type	Parameter(s)	Level	Dur. [ms]	Ramps Type	[ms]	<i>d'</i>
1 ^a	NH	5	18–22, 31	PT	$f = \{250, 500, 700, 800, 900, 1000, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550\}$ Hz	70 dB SPL	500	—	100	1.5
2 ^a	NH	3	25 (av.)	BPN	$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz	85 dB SPL	250	cos	25	1.3
	HI	8	27 (av.)		$bw = \{50, 100, 200, 300, 500\}$ Hz					
3 ^a	NH	2	22, 25	BPN	$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz	80 dB SPL	200	—	25	1.0
	HI	4	20–59		$bw = \{50, 100, 150, 200, 270\}$ Hz					
4 ^a	NH	2	19, 28	BPN	$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz $bw = \{58, 115, 231, 461, 919\}$ Hz	30 dB HL above thr. (Thr. for NH = 0 dB)	300	—	15	1.3
	HI	4	20–59		$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz $bw = \{58, 115, 231, 461, 919\}$ Hz					
5 ^a	NH	10	19–29	BPN	$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz	65 dB SPL	300	linear	15	1.5
	HI	11	20–59		$bw = \{58, 115, 232, 463, 919\}$ Hz					
						thr. < 30 dB: 75 dB SPL 30 dB < thr. < 50 dB: 80 dB SPL thr. > 50 dB: 85 dB SPL				
6 ^a	NH	—	—	BPN	$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz	80 dB SPL	250	cos	25	1.3
	HI	—	—		$bw = \{50, 100, 200, 300, 500\}$ Hz					

^a1: [Brughera et al. \(2013\)](#); 2: [Hawkins and Wightman \(1980\)](#); 3: [Smoski and Trahiotis \(1986\)](#); 4: [Gabriel et al. \(1992\)](#); 5: [Spencer et al. \(2016\)](#); 6: Model exploration of OHC/IHC loss contribution to HI (Sec. III B).

number of required spike trains ranged from 35 up to 85 (per channel). Thus, we used information of 70–170 neurons for each ITD prediction (2 channels, left and right).

The generation of spike trains is a computationally expensive task, thus we wanted to make the most out of the simulated data by creating new spike train combinations across the simulated neurons. Most importantly, we were interested in obtaining distributions of the ITD predictions (Sec. II C) without making any *a priori* assumptions about them (Grün, 2009; Ventura, 2010). Bootstrapping is a suitable technique that helped us achieve both. It consists of generating a *data subset* by randomly sampling a *larger data set* with replacement (Witten and Frank, 2011). In our case, the data subset consisted of the required number of spike trains for each condition (35–85 spike trains per channel, Sec. II A). We defined this number arbitrarily as 20% of the larger data set. Thus, we generated 175–425 spike trains (per channel) in total for each condition.

B. Coincidence detection

1. Shuffled cross correlogram

To quantify the temporal pattern of the AN activity, we used the shuffled cross correlogram (SCC) (Joris *et al.*, 2006; Louage *et al.*, 2004). The SCC is a modified version of the shuffled auto correlogram (SAC) (Joris, 2003). It can be thought as a metric of binaural coincidence detection which compares the timing of the neural firing between the left and right AN fibers. In other words, the SCC quantifies coincident spikes across two different spike trains. The construction of the SCC is illustrated in Fig. 2 and is explained as follows.

First, N spike trains from the left ear and N spike trains from the right ear were used as inputs. Then, the forward and backward time intervals between all the spikes of the first left spike train and all the spikes of all the right spike trains were measured. The same was done with all the spikes of the

FIG. 2. (Color online) Schematic representation of the construction of the SCC (Joris *et al.*, 2006; Louage *et al.*, 2004). Notice how the SCC can be plotted as the number of intervals (or counts) versus the delay value. The parameter $\Delta\tau$ (i.e., the histogram bin width) defines the SCC curve and the model's ITD computation resolution. We chose a value of $20\ \mu\text{s}$.

second left spike train and all the spikes of all the right spike trains and so on. These time intervals were tallied into a histogram with binwidth $\Delta\tau$. The latter can be thought as an integration window: if two spikes occur within this time interval, they are considered to be coincident. Louage *et al.* (2004) originally suggested using a value of $50\ \mu\text{s}$. However, the resolution of the SCC curve—and of the model's ITD predictions in consequence—is limited by it. Since the objective of the model is to predict ITD thresholds that are behaviorally relevant to human perception, $50\ \mu\text{s}$ was too long. We explored using different $\Delta\tau$ values and found a trade-off between SCC resolution and curve smoothness: smaller $\Delta\tau$ values meant better resolution but more jagged SCC curves (and vice versa). Therefore, we decided on a value of $20\ \mu\text{s}$.

We were interested in capturing the firing temporal properties of the AN fibers. Thus, the SCC was normalized by the number of spike trains N , the average firing rate r , $\Delta\tau$, and the stimulus duration D . This was done by dividing the SCC by $N^2 r^2 \Delta\tau D$, making it independent from these parameters and thus dimensionless. In a normalized SCC curve, a count value larger than 1 means that spikes across spike trains tend to be temporally correlated; a count value of 1 shows that there is no (stimulus-induced) temporal correlation; and a count value smaller than 1 indicates anticorrelation (Joris *et al.*, 2006).

Furthermore, we focused our analysis on a window centered at 0 ms spanning $\pm 2\ \text{ms}$ (thus 4 ms in total). This is very short compared to the duration of the stimuli used (Sec. III A). Thus, it was not necessary to correct the SCC curve for the distortion caused by the finite duration of the stimulus. An example SCC curve is shown in Fig. 3(A).

2. Centrality weighting

Different physiological studies of the mammalian auditory system have found that there is a relatively large proportion of neural coincidence detectors that are more sensitive to interaural delays of smaller magnitude (Kuwada *et al.*, 1997; Kuwada *et al.*, 1987; Kuwada and Yin, 1983; Yin *et al.*, 1986). We accounted for this by introducing a centrality weighting function. The latter served two purposes: to reduce the probability of choosing SCC features for the ITD computation that might be ambiguous (Sec. II B 3) and to emphasize the range relevant for human perception.

Several centrality weighting functions have been developed based on psychoacoustic data (e.g., Colburn, 1977; Shackleton *et al.*, 1992; Stern and Shear, 1996). Based on our previous experience, we chose the function proposed by Stern and Shear (1996). This function p depends on the delay τ and is weakly dependent on the stimulus (center) frequency f , as shown in Eq. (1),

$$p(\tau, f) = \begin{cases} 1 & \text{if } |\tau| \leq 200\ \mu\text{s}, \\ \frac{e^{-2\pi k_l(f)|\tau|} - e^{-2\pi k_h|\tau|}}{|\tau|} & \text{if } |\tau| > 200\ \mu\text{s}, \end{cases} \quad (1)$$

with

FIG. 3. (Color online) Different stages of the coincidence detection block for a pure tone with $f=1$ kHz and an imposed ITD of $160 \mu\text{s}$. (A) The raw SCC curve. (B) The centrality weighting function p (adapted from Stern and Shear, 1996). (C) The weighted SCC curve. The ITD was obtained from the latter as the delay value of the maximum peak (dotted line).

$$k_l(f) = \begin{cases} 0.1f^{1.1} & \text{if } f \leq 1200 \text{ Hz,} \\ 0.1(1200)^{1.1} & \text{if } f > 1200 \text{ Hz,} \end{cases}$$

$$k_h = 3000 \text{ s}^{-1}.$$

The generated p spanned across the ± 2 ms duration of the SCC curve [Fig. 3(B)]. The weighted SCC curve was obtained by multiplying p and the original SCC curve element-wise [Fig. 3(C)].

3. Peak selection

We were interested in computing the imposed ITD. We decided to estimate it as the delay value of the maximum peak¹ of the weighted SCC curve, (i.e., the delay value with the largest temporal correlation), as shown in Fig. 3(C).

C. Decision device

For each case, the coincidence detection stage was run 100 times, each time receiving a different set of bootstrapped spike trains (Sec. II A) and yielding as a result a computed ITD value. Then, we generated distributions of the model computations for different ITDs. Said distributions had a time resolution given by the chosen bin size ($20 \mu\text{s}$, Sec. II B 1). In order to make them continuous and smoother, we fitted a cubic spline using the original data points as anchors. The smoothed distributions were scaled to make sure that the area under the curve added up to the number of ITD computed values (100). Example distributions are shown in Fig. 4.

Analogously to a behavioral alternative-forced-choice paradigm, we defined the case where $\text{ITD} = 0$ as the *reference condition* and the case where $\text{ITD} \neq 0$ as the *experimental condition*. For each experiment, we computed the detection index (d') metric (Green and Swets, 1966), given by Eq. (2),

$$d' = \frac{\mu_{\text{ref}} - \mu_{\text{exp}}}{\sqrt{\frac{1}{2}(\sigma_{\text{ref}}^2 + \sigma_{\text{exp}}^2)}}. \quad (2)$$

Then, we built a neurometric function by computing d' across different ITD values (namely, 10, 20, 40, 80, 160, and $320 \mu\text{s}$, Fig. 5). Afterwards, we fitted the sigmoid function given by Eq. (3) across these points,

$$d' = a + \frac{b - a}{1 + 10^{(c - \text{ITD})/d}}, \quad (3)$$

where a is the bottom limit of d' (which we constrained to be not smaller than 0, which translates to a 50% chance of correctly discriminating between the reference and the experimental distribution), b is the top limit of d' (which we constrained to be not larger than 4.65, which translates to a perfect discrimination between the reference and the experimental distribution, Macmillan and Creelman, 2004), c is the ITD_{exp} value that yields a d' value corresponding to 50% of the range of the fitted curve, and d is the slope. The values for these parameters were obtained using a nonlinear regression with iterative least squares estimation. Finally, the ITD threshold was computed as the ITD value corresponding to a point of the fitted sigmoid with a specific d' value. The latter was matched to that of the method used in the behavioral experiment that we wanted to model (Table I). It is worth noting that the predicted ITD threshold is output in ITD units (i.e., μs).

III. MODEL RESULTS

A. Framework validation

We validated the framework's performance in predicting ITD thresholds by comparing its predictions against behavioral data from different studies reported in literature. We were

FIG. 4. (Color online) Computation distributions for a pure tone with $f=1$ kHz with imposed ITD values of $40 \mu\text{s}$ (A), $80 \mu\text{s}$ (B), and $160 \mu\text{s}$ (C).

FIG. 5. (Color online) Neurometric function for a pure tone with $f=1$ kHz. In this case, the ITD threshold was computed as the ITD value corresponding to d' of 1.5 (equivalent to 79% of the psychometric function, as used by Brughera *et al.*, 2013), which resulted in a value of $37.8 \mu\text{s}$.

interested in studies that included NH (Sec. III A 1) and HI (Sec. III A 2) participants and that used narrow-band stimuli (e.g., pure tones, bandpass noise) with ITDs being the only binaural cue present. Table I shows the participants' hearing status, number, and age from the chosen data sets, as well as the characteristics of the stimuli used. In each case, we fed stimuli with these exact same characteristics to the framework. Furthermore, we also generated ITD threshold predictions for additional cases, complementing the behavioral data with computational modelling results. These are shown in Table I in italics.

Unless stated otherwise, we show the geometric mean and the geometric standard deviation (Kirkwood, 1979) as error bars. In all cases, we show the ITD threshold values on a logarithmic scale (Saber, 1995).

1. NH listeners

Figure 6 shows the model's predictions for pure tones in NH listeners. It includes data reported by Brughera *et al.* (2013). For the latter, we show the median of the ITD threshold

for each frequency (together with the original data). When the participants were not able to perform the task, we considered an ITD threshold of $+\infty \mu\text{s}$. We can see that the predicted ITD threshold values follow the trend of the behavioral data.

Figure 7 shows the model's predictions for bandpass noise in NH listeners. It includes data reported by Gabriel *et al.* (1992), Hawkins and Wightman (1980), Smoski and Trahiotis (1986), and Spencer *et al.* (2016). Just like for the pure tones, we can see that overall the framework yields good predictions of the behavioral data. Additionally, the computational modelling predictions suggest that ITD thresholds are the lowest somewhere between 500 and 1000 Hz of the center frequency. Figure 10(A) shows the corresponding dispersion plot. Given the low number of participants in each study, we pooled all data together.

2. HI listeners

One of the most attractive properties of the Zilany *et al.* (2014) and Zilany *et al.* (2009) model is that it is capable of modelling IHC and OHC damage. Physiological studies have revealed that IHC deterioration causes elevation of the AN fiber tuning threshold curves (Liberman and Dodds, 1984), while OHC deterioration additionally causes broadening. Furthermore, OHC damage has also been associated with the reduction in two-tone of AN responses and reduction in the compression of the basilar membrane responses (Liberman, 1984; Liberman and Dodds, 1984; Miller *et al.*, 1997; Robles and Ruggero, 2001; Salvi *et al.*, 1982). The framework incorporates these effects of OHC and IHC impairment using two scaling constants: C_{OHC} and C_{IHC} (Bruce *et al.*, 2003; Zilany and Bruce, 2006), respectively.

C_{OHC} is introduced at the output of the control path. A value of $C_{\text{OHC}}=1$ represents normal OHC function. This allows for normal behavior of the non-linear basilar membrane filter, yielding narrow and low tuning thresholds curves and output compression. The closer C_{OHC} gets to 0, the larger the impairment of the OHCs. This modifies the behavior of the basilar membrane filter in two ways: at low sound levels, it increases the tuning curve bandwidth and

FIG. 6. (Color online) Validation of the modelling framework performance for predicting ITD thresholds for pure tones in NH listeners using the behavioral data reported by Brughera *et al.* (2013) as a reference. (A) Original data and its median compared with model predictions. The latter correspond to one simulated NH listener (and thus have no error bars). The latter were shifted 25 Hz along the x axis for clarity. Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols. The dotted line represents the highest frequency for human ITD discrimination measured by Brughera *et al.* (2013). (B) Dispersion plot of predicted vs behavioral ITD thresholds. The purple line considers data for all frequencies. The blue line considers data for frequencies below 1300 Hz.

FIG. 7. (Color online) Validation of the framework performance for predicting ITD thresholds for bandpass noise in NH listeners using behavioral data reported in different studies as a reference. Error bars denote ± 1 geometric SD across listeners and are reported where available. Model predictions correspond to one simulated NH listener (and thus have no error bars). The latter were shifted 50 Hz along the x axis for clarity. Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols.

elevates the thresholds; at moderate to high levels, it reduces (or eliminates) the output compression.

C_{IHC} is introduced in the signal path. Analogous to C_{OHC} , a C_{IHC} value of 1 represents normal IHC function, while values closer to 0 represent larger IHC impairment. This is modeled by lowering the slope of the function that relates basilar membrane vibration to IHC potential, which causes elevated threshold tuning curves.

The model calculated the C_{OHC} and C_{IHC} values for each listener using as an input his/her reported PTA. It attributed 2/3 of each threshold shift to OHC impairment and the remaining 1/3 to IHC impairment (this proportion is in line with previous studies regarding hearing loss in cats and estimated OHC/IHC detriment in HI listeners in average, Bruce *et al.*, 2003; Lopez-Poveda and Johannesen, 2012; Plack *et al.*, 2004). Specifically, it calculated the listeners' C_{OHC}/C_{IHC} by comparing their OHC/IHC threshold shift with a set of previously computed C_{OHC}/C_{IHC} values based on the measurements by Shera *et al.* (2002). A complete description of their computation is given by Bruce *et al.* (2003) and Zilany and Bruce (2006). Figure 8 shows C_{OHC}/C_{IHC} values as a function of auditory thresholds for different OHC/IHC proportions and for two CFs (500 and 4000 Hz).

FIG. 8. (Color online) C_{OHC} and C_{IHC} values for auditory thresholds for two characteristic frequencies (CF). Different proportions of OHC/IHC impairment (solid/dashed lines) contributing to the threshold shift are shown in different colors.

Figure 9 shows the model's predictions for bandpass noise in HI listeners. It includes data reported by Gabriel *et al.* (1992), Hawkins and Wightman (1980), Smoski and Trahiotis (1986), and Spencer *et al.* (2016). Overall, the framework is capable of following the trends of behavioral data. This is true for data from all studies, except for the case of the Gabriel *et al.* (1992) dataset, where the model is unable to predict an increase of the ITD threshold at 2000 Hz. Additionally, just like in the NH case, the computational modelling predictions suggest that ITD thresholds are the lowest somewhere between 500 and 1000 Hz of the center frequency. It is worth mentioning that, in most cases, the performance of HI listeners shows a large variability, which is reflected by large error bars. Figure 10(B) shows the corresponding dispersion plot. Similarly to the NH case, we pooled all data together.

B. Contribution of OHC/IHC loss to HI

Current literature suggests that although both OHC and IHC impairment are responsible for cochlear hearing loss, OHC damage contributes in a larger proportion to the measured threshold shift on average (Bruce *et al.*, 2003; Moore, 2007; Zilany and Bruce, 2006). However, evidence shows

FIG. 9. (Color online) Validation of the framework performance for predicting ITD thresholds for bandpass noise in HI listeners using behavioral data reported in different studies as a reference. Error bars denote ± 1 geometric SD across listeners and are reported where available. Model predictions correspond to the simulated HI listeners (i.e., $C_{\text{OHC}} < 1$, $C_{\text{IHC}} < 1$). The latter were shifted 100 Hz along the x axis for clarity. Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols.

that this proportion varies broadly individually across listeners, even among those with identical thresholds (Lieberman, 1984; Lopez-Poveda and Johannesen, 2012). Furthermore, anatomical studies have shown that the amount of OHCs and IHCs can vary widely across individuals (McGill and Schuknecht, 1976; Wright *et al.*, 1987). Thus if we wish to individualize predictions, more listener-specific factors may have to be taken into account.

Therefore, we explored the effects of different proportions of hair cell loss on ITD discrimination. Besides the original 2:1 proportion attributed to OHC/IHC (Sec. III A 2), we investigated the cases when the proportion was 1:0, 1:2, and 0:1. We simulated two HI listeners with a mild and a moderate impairment with flat threshold shifts of 25 and 40 dB hearing level (HL), respectively, across all frequencies (plus a NH

listener as a reference). As a stimulus, we used bandpass noise with characteristics defined in the last row of Table I.

These results are shown in Fig. 11. On one hand, we can see that in the 25 dB HL case (panel A) there is a slight increase of the thresholds compared to the NH case. However, the curves show little spread, making it hard to see the effect of different OHC/IHC proportions. On the other hand, we can see that in the 40 dB HL case (panel B) not only are the thresholds higher, but the curves have a larger spread. These curves suggest that there is an effect of the different OHC/IHC proportions for frequencies above 500 Hz. Furthermore, it looks like the impairment of OHC has a larger impact on ITD detection: when the auditory threshold shift is completely attributed to the OHCs, the ITD thresholds are the worst (i.e., highest).

FIG. 10. (Color online) Dispersion plots of predicted vs behavioral ITD thresholds for bandpass noise stimuli. Data points from all studies [■ Hawkins and Wightman (1980), ◆ Smoski and Trahiotis (1986), ▲ Gabriel *et al.* (1992), ▼ Spencer *et al.* (2016)] were pooled together. In the NH case (A), the framework tends to overestimate the ITD thresholds, while in the HI case (B), the framework tends to underestimate them.

FIG. 11. (Color online) Effect of attributing different auditory threshold shift proportions to OHC/IHC impairment for bandpass noise. Its corresponding parameters are given in the last row of Table I.

IV. DISCUSSION

In this work, we introduced a computational framework that uses a physiologically inspired model of the AN as a front end and a neurometric decision device as a back end to predict ITD thresholds in NH and HI listeners. We validated its performance by comparing the predicted ITD thresholds for narrow-band stimuli against behavioral data reported in literature. Additionally, we investigated the effect of changing the proportion of impairment attributed to different OHC/IHC combinations.

A. Framework structure

The presented computational framework has several advantages. First, the model of the auditory periphery used as a front end (Zilany *et al.*, 2014; Zilany *et al.*, 2009, Sec. II A) is inspired by the anatomy and physiology of the AN and has been validated with a wide range of physiological data. Furthermore, being able to model the individual elements of the auditory pathway allowed us to study them individually. In this particular case, it allowed us to incorporate sensorineural hearing loss as impairment of either the OHCs or the IHCs, something which would not be possible if the representation of the AN was coarser.

Regarding the coincidence detection stage, the SCC (Joris *et al.*, 2006; Louage *et al.*, 2004) basically consists of tallying spike intervals. This process can be thought as a natural display of the information arriving from the auditory periphery (Joris *et al.*, 2005). In other words, it extracts information by comparing temporal coding across contralateral AN fibres within an integration window. Our results show that such a relatively simple metric is enough to model ITD discrimination. Additionally, the SCC helped us transform a discrete representation of neural activity (i.e., spike trains) into a continuous representation (i.e., the SCC curve itself), allowing us to further manipulate said representation accordingly (centrality weighting) in order to compute an ITD value (peak selection).

Finally, the implemented neurometric decision device used the distribution of these computations to predict an ITD threshold value. It is worth emphasizing that the latter was given in time units (μ s in this case), which made the comparison between the model predictions and the behavioral data intuitive and straightforward. Using the d' metric allowed us to incorporate information about the framework's ITD computations and their dispersion (i.e., variability).

Furthermore, the framework's modular design allowed the use of different building blocks at any of its different stages. For instance, the current auditory periphery block could be substituted by any other model that takes a pressure wave as an input and yields spike trains as an output. This is not restricted to models of purely the AN. Models of electrical stimulation could be used to investigate binaural hearing in bilateral cochlear implant users (e.g., Prokopiou *et al.*, 2017).

During the framework's development, we were careful to comply with the suggestions proposed by Colburn and Durlach (1978).

- (1) Relate the model's assumptions and parameters with known physiological data. Our framework included a front end that mimics the response of the AN based on physiological information.
- (2) Avoid overfitting of the model's predictions due to excessive parametrization. The framework's parameters were not tuned in a per-case basis. They were defined *a priori* and were the same for all conditions of all simulated experiments [except for the decision device criterion (Sec. II C), which was configured according to the simulated behavioral experiment]. Furthermore, we kept the number of parameters to a minimum.
- (3) Describe quantitatively how the stimuli are processed and how they are affected by (internal) noise. The modular nature of our approach allowed us to examine not only the final output, but also the outputs of the intermediate stages (Sec. II). Additionally, the framework's internal noise (e.g., from the auditory periphery, from the coincidence detection) was taken into account by the decision device to predict the ITD thresholds.
- (4) Consider perceptual principles when modelling higher, more central portions of the system. The implemented decision device (Sec. II C) models psychophysical aspects of human perception (Green and Swets, 1966).
- (5) Compare the model predictions with relevant data. We validated the framework with appropriate behavioral datasets previously reported in literature (Sec. III A).

B. Framework validation

The proposed framework was able to predict the trends of behavioral data. In the NH case, we predicted ITD thresholds for pure tones and bandpass noise. For the former, we used the dataset from Brughera *et al.* (2013) as a reference. Figure 6(A) shows that although the framework tended to overestimate the

ITD thresholds, it was still able to capture the data trend: low thresholds in the mid-frequency range (700–1000 Hz) and higher thresholds in the low (250–500 Hz) and high (1200–1400 Hz) frequency ranges (with a considerable non-monotonicity at ~1300 Hz). However, the model does not increase its ITD threshold output as fast as the behavioral data. Additional computational simulations show that the model is only unable to “perform the task” (i.e., output ITD threshold values that were too large) above 1500 Hz. We hypothesized that this was because the auditory periphery stage is unable to capture the loss of phase locking for frequencies above 1300 Hz fast enough. We confirmed this after computing the vector strength (Johnson, 1980) for the spike trains across different frequencies. We obtained vector strength values that slowly decreased monotonically with increasing frequency with an abrupt decrease only at ~1500 Hz. Therefore, the model finds its performance limit at higher frequencies than we would expect.

In the case of bandpass noise, we used different datasets (Fig. 7). The predictions of the datasets from Smoski and Trahiotis (1986) and Spencer *et al.* (2016) were very close to the reported mean values. The predictions of the datasets from Hawkins and Wightman (1980) and Gabriel *et al.* (1992)² were also able to follow the data trend, although thresholds were overestimated. The model overestimation of the NH thresholds was caused by the SCC resolution. We chose a bin size of 20 μ s (smaller values yielded noisy curves leading to incorrect predictions and larger values were too far from human performance). Therefore, the distributions of short experimental ITDs (e.g., 10 or 20 μ s) overlapped largely with the distributions of the reference condition (ITD = 0 μ s), yielding low d' values. After analysis of different neurometric curves, we found that the points corresponding to these short ITDs were responsible for decreasing the slope of the sigmoid fit of the neurometric function, causing the decision device to output higher ITD thresholds. Additionally, the modelling predictions revealed an underlying function that suggests that the best performance (i.e., lowest thresholds) is achieved when the bandpass noise center frequency is between 500 and 1000 Hz. However, this would need to be confirmed with further behavioral measurements.

It is worth mentioning that the model is not able to fully account for the performance detriment between 1000 and 2000 Hz of the Gabriel *et al.* (1992) data (which is the only study that includes behavioral data for NH listeners at that frequency). The model shows increased thresholds between these two frequencies, but this increase is not as large as in the behavioral case. We attribute this to the model still being sensitive to the temporal fine structure of the stimulus at 2000 Hz. We investigated this by computing the difcor (Louage *et al.*, 2004). The peak of the difcor reflects the temporal fine structure coding strength of the spike trains (and thus of the model). First, we computed the SAC for the left ear spike trains. Then, we computed the SCC between the left and the right ear spike trains. However, the latter corresponded to a polarity-inverted version of the stimulus. Finally, we subtracted them bin by bin. We found that at 1000 Hz, there was strong temporal fine structure coding,

with a peak of 3.44. This value decreased only to 2.69 at 2000 Hz. A larger loss of synchrony would result in larger (and thus more accurate) predicted thresholds, as it is in the case at 4000 Hz (with a peak difcor value close to 0).

In the HI case, we predicted ITD thresholds for bandpass noise from the same datasets as in the NH case (Gabriel *et al.*, 1992; Hawkins and Wightman, 1980; Smoski and Trahiotis, 1986; Spencer *et al.*, 2016). Figure 9 shows that the framework was able to predict the behavioral trend at a group level. Low center frequency values yielded low ITD thresholds, which the model was able to predict accurately. Increasing center frequency values yielded higher ITD thresholds. The framework was able to capture this trend, although its predictions tended to be lower than the actual behavioral data (in contrast to the NH case). However, these results have to be handled with care, since we cannot affirm that group (i.e., mean) results can be translated to each individual participant (Akeroyd, 2014). High intersubject variability is a common issue in most HI studies, including the ones presented here. This variability could be attributed to a variety of reasons, being low correlations between PTA and listeners' performance in ITD threshold detection tasks one that is commonly suggested (Moore, 2007). We used PTA information to model OHC/IHC impairment (i.e., to compute C_{OHC} and C_{IHC} values, Sec. III A 2). We believe this is a good first step towards incorporating HI in our framework. Additionally, just like in the NH case, the modelling predictions suggest that lowest thresholds are achieved when the bandpass noise center frequency is between 500 and 1000 Hz. Likewise, this would need to be confirmed with further behavioral measurements (similar to those performed by Gabriel *et al.*, 1992).

C. Contribution of OHC/IHC loss to HI

In the case of the simulated listener with mild impairment [25 dB HL, Fig. 11(A)], we saw that the thresholds were not so different from those of a NH listener. This was expected since thresholds of 25 dB HL are just below the conventional definition of NH of 20 dB HL across all frequencies. Additionally, the stimulus level was well above said threshold.

Of more interest is the case of the simulated listener with moderate impairment [40 dB HL, Fig. 11(B)]. These results hint that the impairment of OHCs has a larger impact on the thresholds than impairment of the IHCs. OHC impairment is modelled by the Zilany *et al.* (2014) and Zilany *et al.* (2009) front end as loss of compression, broader tuning, and elevated thresholds (i.e., decreased gain) of the control path (Zilany and Bruce, 2006). These changes in the input-output function as well as in the bandwidth are responsible for the detriment of the temporal coding of the neural fibers. Furthermore, the compression parameters of the model are frequency-dependent (Zhang *et al.*, 2001). The filters' responses are almost linear for frequencies below 500 Hz, while having compressing nonlinearities for frequencies above 500 Hz. This explains why we do not see an effect of different OHC/IHC proportions at lower frequencies.

Lopez-Poveda and Johannesen (2012) showed that the OHC/IHC dysfunction contribution to hearing impairment can vary largely across cases, even in listeners with comparable audiometric losses. These results together with the presented model simulations could partially explain the variability observed in the performance of HI listeners in ITD discrimination tasks.

D. Comparison with existing models

Ideally, we would like to quantitatively compare our framework's performance with other similar existing models. Unfortunately, doing such a systematic analysis is not a trivial task (e.g., Ashida *et al.*, 2017; Saremi *et al.*, 2016) and would require a more homogeneous benchmark (Dietz *et al.*, 2018), which is not available for all models. Therefore, further discussion uses a qualitative approach.

First of all, it is worth mentioning that the idea of including physiologically inspired concepts as components of larger models is not new. For example, Patterson *et al.* (1995) used the IHC model proposed by Meddis (1988) (which simulates neurotransmitter flow across reservoirs) to convert the basilar membrane motion (coming from either a gammatone or a transmission line filter) into a neural activity pattern. The Zilany *et al.* (2014) and Zilany *et al.* (2009) model has been used as a front end for studying the effect of sensorineural hearing loss on speech intelligibility (Heinz, 2015) and on sound localization in the median plane (Baumgartner *et al.*, 2016). Moreover, very recent work has used that same front end to study ILD perception. The model proposed by Brown and Tollin (2016) subtracted simulated left and right ear AN spike trains within a running temporal window. This simple model was good enough to validate their physiological and psychophysical observations of ILD sensitivity and robustness. Laback *et al.* (2017) developed a more elaborate framework. They used the simulated spike trains as an input to an interaural comparison stage where they evaluated the difference in mean discharge rates between left and right ear AN inputs and the variability of these rates over different stimuli presentations. Their model was able to account for a variety of published ILD perception data as well as their own.

Nonetheless, few models have used a physiologically inspired architecture to study ITD discrimination. To start with, in our previous work (Prokopiou *et al.*, 2017) we investigated binaural temporal discrimination of NH and bilateral cochlear implant users with different stimuli using physiologically based front ends. However, we did not account for the large proportion of neurons sensitive to smaller ITDs (incorporated in the presented framework by using a central-ity weighting function, Sec. II B 2). The ITD threshold estimation was done using a novel binary classifier characterisation device. Although the latter yielded good data trend predictions, its approach was more analytical (rather than directly modelling psychophysical procedures). Additionally, the model predictions were given in arbitrary model units (rather than in proper time units [μ s], like the current framework), which made the comparison with

behavioral data more difficult. Furthermore, we did not investigate the (unaided) HI case.

Hancock and Delgutte (2004) used a neural-pooling model to simulate ITD-sensitive IC neurons to broadband noise. They used gammatone filters to model the processing of the auditory periphery all the way to the IC. Then, they generated a population model which they parametrized with their own physiological data collected in cats. At a single-neuron level, they found very good agreement between physiological data and model predictions of rate-ITD curves. At a population level, they were able to predict ITD thresholds for pure tones and broadband noise. However, using a filter-bank to simulate the auditory periphery does not include important details of its physiology and limits its capabilities in further expanding it to include HI at this level. Additionally, they included a so-called “efficiency parameter” (ϵ) for the ITD thresholds predictions. This parameter was empirically adjusted to account for several factors such as stimulus variability, inefficient pooling process, or deficient decision making, but had no physiological relevance. They acknowledged that this approach gave the absolute ITD thresholds predicted values little significance (the comparison with behavioral data was not so straightforward) and thus focused in the trends of ITD threshold changes.

Brughera *et al.* (2013) minutely investigated ITD threshold discrimination in NH listeners using pure tones across different frequencies, with special focus in the range between 1200 and 1400 Hz. Additionally, they predicted this data set using two different variations of a Hodgkin-Huxley-based MSO neuron model: a lateralization centroid model and a rate-difference model. The former was successful in predicting ITD thresholds at high frequencies, while the latter was successful in predicting ITD thresholds at low and middle frequencies. This rate-difference variation could also predict high frequency thresholds, but only when the model was tuned *ad hoc*. They also proposed using a combination of both types of models in order to predict thresholds across all frequencies. However, neither of these two scenarios (stimulus-specific parameter tuning and/or processing) are desired, since they could very well lead to overfitting.

E. Framework limitations

Finally, even though the presented framework was successful in predicting ITD threshold perception trends in NH and HI listeners, there are some shortcomings worth pointing out.

We wanted to investigate the contributions of physiological pre-processing stages shaping the input to the binaural stage. Our results show that using a physiologically based model of the auditory periphery combined with a coincidence detection stage (such as the SCC) together with a neurometric decision device can be sufficient to model simple binaural processes. In other words, combining information from the auditory periphery with a simple metric of binaural coincidence is enough to model ITD discrimination in NH and HI listeners. Other studies have used a similar approach. For instance, Franken *et al.* (2014) input cat AN and trapezoid body neural recordings to a “bare bones” model of an

MSO neuron. This neuron generated an output spike when its inputs occurred close enough in time. They concluded that the fundamental operation in the mammalian binaural circuit is the coincidence counting of (single) binaural input spikes. This type of simple binaural schemes (either based on metrics or on models) go in line with recent physiological studies that suggest that the ITD sensitivity depends on the exact timing of the excitatory inputs to MSO neurons (van der Heijden *et al.*, 2013). Including a more elaborate MSO stage [such as the models proposed by Brughera *et al.* (2013), Colburn *et al.* (2009), Colburn *et al.* (1990), and Takanen *et al.* (2014)] could still be beneficial. For example, the MSO has been associated with processing of temporal fine structure, while the lateral superior olive has been associated with processing of the envelope (Remme *et al.*, 2014). Including a two-channel structure (similar to the one used by Dietz *et al.*, 2009) could allow us to further investigate the effect of HI in temporal fine structure and envelope binaural perception. It could even help reduce the discrepancies between the model predictions and the behavioral data of high-frequency pure tones, since it has been suggested that synchrony loss takes place there (Brughera *et al.*, 2013). Additionally, modelling the MSO neural activity in more detail could help us discern its contribution to auditory brain stem responses (Ashida *et al.*, 2015). Finally, an MSO stage would be crucial to understand better the processing of (sensorineurally impaired) peripheral input along the binaural auditory pathway.

In the centrality weighting block (Sec. II B 2) of the coincidence detection stage, we used the function proposed by Stern and Shear (1996). This function was obtained using human psychoacoustical data. It would be interesting to explore the effect of centrality-weighting functions obtained using physiological data (e.g., Hancock and Delgutte, 2004; McAlpine *et al.*, 2001). Although these functions were generated using animal data, their use could help in investigating if similar neural distributions exist in humans and to provide a better insight on the underlying assumptions of the psychoacoustic functions.

For the sake of simplicity, we tuned the AN fibers with the same CF as the stimulus frequency (for pure tones) or center frequency (for bandpass noise). Tuning AN fibers to different CFs could allow investigating the effect of cross-frequency integration for ITD perception in NH and HI listeners for broadband stimuli. It could also allow to explore the contribution of off-frequency filters to ITD sensitivity of modulated stimuli (Bernstein and Trahiotis, 2002).

Additionally, we considered ITDs to be the only binaural cue present. However, this is unrealistic, since sounds in a real-world scenario contain ITDs together with ILDs. In order to be able to model this, we would need an additional module that could integrate binaural information across cues. The latter is not a trivial task, since the questions of at what levels of the auditory pathway and to what extent are time and level cues combined are still topics of ongoing research (Brown and Tollin, 2016; Ellinger *et al.*, 2017; Johnson and Hautus, 2010; Palomäki *et al.*, 2005; Phillips and Hall, 2005; Takanen *et al.*, 2014).

Furthermore, the framework fully attributed HI to hair cell damage (Sec. III A 2). However, there are additional factors that may affect binaural perception, such as age, cognition, and loss of AN-hair cell synapses (Gallun *et al.*, 2014; Kujawa and Liberman, 2015).

Last, it is worth mentioning that several assumptions on which the framework's blocks rely on have been validated using animal data. Their translation to human auditory perception still needs to be confirmed by further physiological and psychoacoustical studies (e.g., Salminen *et al.*, 2018).

V. CONCLUSIONS

We presented a physiologically based modelling framework capable of predicting ITD thresholds for NH and HI listeners. It uses the model of the AN proposed by Zilany *et al.* (2014) and Zilany *et al.* (2009) as a front end and a coincidence detection stage based on the SCC (Joris *et al.*, 2006; Louage *et al.*, 2004) together with a neurometric decision device as a back end. The framework was validated by comparing its predictions with behavioral data from literature, showing good agreement between them. These results show that the presented framework is capable of modelling ITD discrimination of NH and HI listeners at a group level. Additionally, we used it to study the contribution of OHC/IHC loss of HI listeners. Model results hint that OHC impairment has a larger impact on ITD discrimination thresholds than damage to the IHCs for frequencies >500 Hz.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA Grant Agreement No. 317521 (ICanHear) and from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 637424 (ERC Starting Grant to T.F.). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation–Flanders (FWO) and the Flemish Government Department EWI. We would like to thank the editor Dr. Mathias Dietz and the two anonymous reviewers for their helpful comments and feedback. Last, we would also like to thank Dr. Christoph Scheidiger and Professor Ian C. Bruce for the fruitful discussions.

¹In their models, Stern and Colburn (1978) and Stern and Shear (1996) estimated the ITD as the centroid of the number of observed coincidences along the delay axis. Bernstein and Trahiotis (2011) found that using the centroid as the decision variable (instead of the maximum peak) was a key factor in their model. However, in our case we explored using the centroid of the SCC curve and reached a different conclusion. Quite often, the SCC curve had a clear peak (which reflected important coincident activity). We found that when using the centroid, this information was lost. The computed ITD distributions of the target cases overlapped quite a lot with the reference one. Additionally, they were much broader (compared to those obtained using the maximum peak). Both of these factors contributed to yielding lower d' values, therefore reducing the slope of the fitted sigmoid and predicting ITD thresholds that were between 200 and 300 μ s higher than the behavioral ones. Furthermore, when using the centroid, the model

was unable to predict the frequency region of best performance. Therefore, computing the maximum peak was a more suitable choice for our application.

²This study included only two listeners. However, only the mean was reported, thus we were not able to report the individual data.

- Akeroyd, M. A. (2014). "An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners," *Trends Hear.* **18**, 1–7.
- Akeroyd, M. A., and Whitmer, W. M. (2016). "Spatial hearing and hearing aids," in *Hearing Aids* (Springer, Berlin), Chap. 7, pp. 181–215.
- Ashida, G., Funabiki, K., and Kretzberg, J. (2015). "Minimal conductance-based model of auditory coincidence detector neurons," *PLoS One* **10**(4), e0122796.
- Ashida, G., Tollin, D. J., and Kretzberg, J. (2017). "Physiological models of the lateral superior olive," *PLoS Comput. Biol.* **13**(12), e1005903.
- 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. (2001). "Auditory processing of interaural timing information: New insights," *J. Neurosci. Res.* **66**(6), 1035–1046.
- Bernstein, L. R., and Trahiotis, C. (2002). "Enhancing sensitivity to interaural delays at high frequencies by using 'transposed stimuli,'" *J. Acoust. Soc. Am.* **112**(3), 1026–1036.
- Bernstein, L. R., and Trahiotis, C. (2011). "Lateralization produced by envelope-based interaural temporal disparities of high-frequency, raised-sine stimuli: Empirical data and modeling," *J. Acoust. Soc. Am.* **129**(3), 1501–1508.
- Bernstein, L. R., and Trahiotis, C. (2012). "Lateralization produced by interaural temporal and intensive disparities of high-frequency, raised-sine stimuli: Data and modeling," *J. Acoust. Soc. Am.* **131**(1), 409–415.
- Bidelman, G. M., and Heinz, M. G. (2011). "Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing," *J. Acoust. Soc. Am.* **130**(3), 1488–1502.
- Blauert, J. (1997). *Spatial Hearing—The Psychophysics of Human Sound Localization* (The MIT Press, Cambridge, MA).
- 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**(38), 9908–9921.
- Bruce, I. C., Léger, A. C., Moore, B. C., and Lorenzi, C. (2013). "Physiological prediction of masking release for normal-hearing and hearing-impaired listeners," in *Proceedings of Meetings on Acoustics*, Acoustical Society of America, Montreal, Canada, Vol. 19, pp. 1–8.
- Bruce, I. C., Sachs, M. B., and Young, E. D. (2003). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," *J. Acoust. Soc. Am.* **113**(1), 369–388.
- 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**(5), 2839–2855.
- Colburn, H., and Latimer, J. (1978). "Theory of binaural interaction based on auditory-nerve data. III. Joint dependence on interaural time and amplitude differences in discrimination and detection," *J. Acoust. Soc. Am.* **64**(1), 95–106.
- Colburn, H. S. (1973). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," *J. Acoust. Soc. Am.* **54**(6), 1458–1470.
- Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**(2), 525–533.
- Colburn, H. S., Chung, Y., Zhou, Y., and Brughera, A. (2009). "Models of brainstem responses to bilateral electrical stimulation," *J. Assoc. Res. Otolaryngol.* **10**(1), 91–110.
- Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction," in *Handbook of Perception, IV (Hearing)*, edited by E. Carterette and M. Friedman (Academic Press, Colburn and Durlach, New York), pp. 467–518.
- Colburn, H. S., Shinn-Cunningham, B., Kidd, G., Jr., and Durlach, N. (2006). "The perceptual consequences of binaural hearing: Las consecuencias perceptuales de la audición binaural," *Int. J. Audiol.* **45**, 34–44.
- Colburn, H. S., Yan-an, H., and Culotta, C. P. (1990). "Coincidence model of MSO responses," *Hear. Res.* **49**(1–3), 335–346.
- Dietz, M., Ewert, S. D., and Hohmann, V. (2009). "Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities," *J. Acoust. Soc. Am.* **125**(3), 1622–1635.
- Dietz, M., Lestang, J.-H., Majdak, P., Stern, R. M., Marquardt, T., Ewert, S. D., Hartmann, W. M., and Goodman, D. F. (2018). "A framework for testing and comparing binaural models," *Hear. Res.* **360**, 92–106.
- Durlach, N., Thompson, C., and Colburn, H. (1981). "Binaural interaction in impaired listeners: A review of past research," *Audiology* **20**(3), 181–211.
- Ellinger, R. L., Jakien, K. M., and Gallun, F. J. (2017). "The role of interaural differences on speech intelligibility in complex multi-talker environments," *J. Acoust. Soc. Am.* **141**(2), EL170–EL176.
- Franken, T. P., Bremen, P., and Joris, P. X. (2014). "Coincidence detection in the medial superior olive: Mechanistic implications of an analysis of input spiking patterns," *Front. Neural Circ.* **8**, 42.
- Gabriel, K. J., Koehnke, J., and Colburn, H. S. (1992). "Frequency dependence of binaural performance in listeners with impaired binaural hearing," *J. Acoust. Soc. Am.* **91**(1), 336–347.
- 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**(4), 802–813.
- Gallun, F. J., Mcmillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M., and Konrad-Martin, D. (2014). "Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity," *Audit. Cogn. Neurosci.* **8**, 172.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).
- Grothe, B., Pecka, M., and McAlpine, D. (2010). "Mechanisms of sound localization in mammals," *Physiol. Rev.* **90**(3), 983–1012.
- Grün, S. (2009). "Data-driven significance estimation for precise spike correlation," *J. Neurophysiol.* **101**(3), 1126–1140.
- Hancock, K. E., and Delgutte, B. (2004). "A physiologically based model of interaural time difference discrimination," *J. Neurosci.* **24**(32), 7110–7117.
- Hawkins, D. B., and Wightman, F. L. (1980). "Interaural time discrimination ability of listeners with sensorineural hearing loss," *Audiology* **19**(6), 495–507.
- Heinz, M. G. (2015). "Neural modelling to relate individual differences in physiological and perceptual responses with sensorineural hearing loss," in *Proceedings of the International Symposium on Auditory and Audiological Research*, Vol. 5, pp. 137–148.
- Heinz, M. G., and Swaminathan, J. (2009). "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech," *J. Assoc. Res. Otolaryngol.* **10**(3), 407–423.
- Jeffress, L. A. (1948). "A place theory of sound localization," *J. Compar. Physiol. Psychol.* **41**(1), 35–39.
- Jennings, S. G., Heinz, M. G., and Strickland, E. A. (2011). "Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot," *J. Assoc. Res. Otolaryngol.* **12**(3), 345–360.
- Jennings, S. G., and Strickland, E. A. (2012). "Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity," *J. Acoust. Soc. Am.* **132**(4), 2483–2496.
- Johnson, B. W., and Hautus, M. J. (2010). "Processing of binaural spatial information in human auditory cortex: Neuromagnetic responses to interaural timing and level differences," *Neuropsychologia* **48**(9), 2610–2619.
- 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**(4), 1115–1122.
- Joris, P. X. (2003). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," *J. Neurosci.* **23**(15), 6345–6350.
- Joris, P. X., Louage, D. H., Cardoen, L., and van der Heijden, M. (2006). "Correlation index: A new metric to quantify temporal coding," *Hear. Res.* **216–217**, 19–30.
- Joris, P. X., van der Heijden, M., Louage, D. H., Van de Sande, B., and Van Kerckhoven, C. (2005). "Dependence of binaural and cochlear 'best delays' on characteristic frequency," in *Auditory Signal Processing* (Springer, Berlin), pp. 477–483.
- King, A., Hopkins, K., and Plack, C. J. (2014). "The effects of age and hearing loss on interaural phase difference discrimination," *J. Acoust. Soc. Am.* **135**(1), 342–351.
- Kirkwood, T. (1979). "Geometric mean and measures of dispersion," *Biometrics* **19**, 908–909.
- Kujawa, S. G., and Liberman, M. C. (2015). "Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss," *Hear. Res.* **330**, 191–199.
- Kuwada, S., Batra, R., and Fitzpatrick, D. C. (1997). "Neural processing of binaural temporal cues," in *Binaural and Spatial Hearing in Real and Virtual Environments* (Lawrence Erlbaum, Mahwah, NJ), pp. 399–425.

- Kuwada, S., Stanford, T. R., and Batra, R. (1987). "Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: Effects of changing frequency," *J. Neurophysiol.* **57**(5), 1338–1360.
- Kuwada, S., and Yin, T. C. (1983). "Binaural interaction in low-frequency neurons in inferior colliculus of the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function," *J. Neurophysiol.* **50**(4), 981–999.
- Laback, B., Dietz, M., and Joris, P. (2017). "Temporal effects in interaural and sequential level difference perception," *J. Acoust. Soc. Am.* **142**(5), 3267–3283.
- Liberman, M. C. (1984). "Single-neuron labeling and chronic cochlear pathology. I. Threshold shift and characteristic-frequency shift," *Hear. Res.* **16**(1), 33–41.
- Liberman, M. C., and Dodds, L. W. (1984). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," *Hear. Res.* **16**(1), 55–74.
- 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**(6), 1608–1622.
- Lopez-Poveda, E. A., and Johannesen, P. T. (2012). "Behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss," *J. Assoc. Res. Otolaryngol.* **13**(4), 485–504.
- Louage, D. H., Joris, P. X., and van der Heijden, M. (2006). "Decorrelation sensitivity of auditory nerve and anteroventral cochlear nucleus fibers to broadband and narrowband noise," *J. Neurosci.* **26**(1), 96–108.
- Louage, D. H., van der Heijden, M., and Joris, P. X. (2004). "Temporal properties of responses to broadband noise in the auditory nerve," *J. Neurophysiol.* **91**(5), 2051–2065.
- Macmillan, N. A., and Creelman, C. D. (2004). *Detection Theory: A User's Guide* (Psychology Press, Hove, East Sussex).
- McAlpine, D. (2005). "Creating a sense of auditory space," *J. Physiol.* **566**(1), 21–28.
- McAlpine, D., Jiang, D., and Palmer, A. R. (2001). "A neural code for low-frequency sound localization in mammals," *Nat. Neurosci.* **4**(4), 396–401.
- McGill, T. J., and Schuknecht, H. F. (1976). "Human cochlear changes in noise induced hearing loss," *Laryngoscope* **86**(9), 1293–1302.
- Meddis, R. (1988). "Simulation of auditory–neural transduction: Further studies," *J. Acoust. Soc. Am.* **83**(3), 1056–1063.
- Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (1997). "Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers," *J. Acoust. Soc. Am.* **101**(6), 3602–3616.
- Moncada-Torres, A., van Wieringen, A., Bruce, I. C., Wouters, J., and Francart, T. (2017). "Predicting phoneme and word recognition in noise using a computational model of the auditory periphery," *J. Acoust. Soc. Am.* **141**(1), 300–312.
- Moore, B. C. (1996). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," *Ear Hear.* **17**(2), 133–161.
- Moore, B. C. J. (2007). *Cochlear Hearing Loss: Physiological, Psychological, and Technical Issues* (Wiley Series in Human Communication Science, Chichester).
- Palomäki, K. J., Tiitinen, H., Mäkinen, V., May, P. J., and Alku, P. (2005). "Spatial processing in human auditory cortex: The effects of 3D, ITD, and ILD stimulation techniques," *Cogn. Brain Res.* **24**(3), 364–379.
- Patterson, R. D., Allerhand, M. H., and Giguere, C. (1995). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.* **98**(4), 1890–1894.
- Peelle, J. E., and Wingfield, A. (2016). "The neural consequences of age-related hearing loss," *Trends Neurosci.* **39**(7), 486–497.
- Phillips, D. P., and Hall, S. E. (2005). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," *Hear. Res.* **202**(1), 188–199.
- Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (2004). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," *J. Acoust. Soc. Am.* **115**(4), 1684–1695.
- Prokopiou, A., Moncada-Torres, A., Wouters, J., and Francart, T. (2017). "Functional modelling of interaural time difference discrimination in acoustical and electrical hearing," *J. Neural Eng.* **14**(4), 046021.
- Remme, M. W., Donato, R., Mikiel-Hunter, J., Ballesterio, J. A., Foster, S., Rinzel, J., and McAlpine, D. (2014). "Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues," *Proc. Natl. Acad. Sci.* **111**(22), E2339–E2348.
- Robles, L., and Ruggero, M. A. (2001). "Mechanics of the mammalian cochlea," *Physiol. Rev.* **81**(3), 1305–1352.
- Saberi, K. (1995). "Some considerations on the use of adaptive methods for estimating interaural-delay thresholds," *J. Acoust. Soc. Am.* **98**(3), 1803–1806.
- Salminen, N. H., Jones, S. J., Christianson, G. B., Marquardt, T., and McAlpine, D. (2018). "A common periodic representation of interaural time differences in mammalian cortex," *NeuroImage* **167**, 95–103.
- Salvi, R., Perry, J., Hamernik, R. P., and Henderson, D. (1982). "Relationships between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma," in *New Perspectives on Noise-Induced Hearing Loss* (Raven, New York), pp. 165–188.
- Saremi, A., Beutelmann, R., Dietz, M., Ashida, G., Kretzberg, J., and Verhulst, S. (2016). "A comparative study of seven human cochlear filter models," *J. Acoust. Soc. Am.* **140**(3), 1618–1634.
- Sayers, B. M., and Cherry, E. C. (1957). "Mechanism of binaural fusion in the hearing of speech," *J. Acoust. Soc. Am.* **29**(9), 973–987.
- Shackleton, T. M., Meddis, R., and Hewitt, M. J. (1992). "Across frequency integration in a model of lateralization," *J. Acoust. Soc. Am.* **91**(4), 2276–2279.
- 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.* **99**(5), 3318–3323.
- Smoski, W. J., and Trahiotis, C. (1986). "Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss," *J. Acoust. Soc. Am.* **79**(5), 1541–1547.
- Spencer, N. J., Hawley, M. L., and Colburn, H. S. (2016). "Relating interaural difference sensitivities for several parameters measured in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **140**(3), 1783–1799.
- Stecker, G. C., and Gallun, F. J. (2012). "Binaural hearing, sound localization, and spatial hearing," in *Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing* (Plural, San Diego, CA), Chap. 14, pp. 383–433.
- Stern, R. M., and Colburn, H. S. (1978). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," *J. Acoust. Soc. Am.* **64**(1), 127–140.
- Stern, R. M., and Shear, G. D. (1996). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," *J. Acoust. Soc. Am.* **100**(4), 2278–2288.
- Stern, R. M., and Trahiotis, C. (1992). "The role of consistency of interaural timing over frequency in binaural lateralization," in *Auditory Physiology and Perception* (Elsevier, Amsterdam), pp. 547–554.
- Takanen, M., Santala, O., and Pulkki, V. (2014). "Visualization of functional count-comparison-based binaural auditory model output," *Hear. Res.* **309**, 147–163.
- van der Heijden, M., Lorteije, J. A., Plauška, A., Roberts, M. T., Golding, N. L., and Borst, J. G. G. (2013). "Directional hearing by linear summation of binaural inputs at the medial superior olive," *Neuron* **78**(5), 936–948.
- Ventura, V. (2010). "Bootstrap tests of hypotheses," in *Analysis of Parallel Spike Trains*, edited by S. Grün and S. Rotter (Springer, Berlin), Vol. 7, Chap. 18, pp. 383–398.
- Wang, D., and Brown, G. J. (2005). "Binaural sound localization," in *Computational Auditory Scene Analysis* (Wiley, New York), Chap. 5, pp. 1–34.
- 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**(1), 164–181.
- Witten, I., and Frank, E. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. (Morgan Kaufmann, Burlington, MA).
- Wright, A., Davis, A., Bredberg, G., Ulehlova, L., and Spencer, H. (1987). "Hair cell distributions in the normal human cochlea," *Acta Otolaryngol. Suppl.* **444**, 1–48.
- Yin, T. C., Chan, J. C., and Irvine, D. R. (1986). "Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. I. Responses to wideband noise," *J. Neurophysiol.* **55**(2), 280–300.
- 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**(2), 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**(3), 1446–1466.

- 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**(1), 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**(5), 2390–2412.
- Zilany, M. S., and Carney, L. H. (2010). "Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics," *J. Neurosci.* **30**(31), 10380–10390.