PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Integrating interaural differences of time and level across frequencies and with each other in a precedence effect model

M. Torben PASTORE[(1)], Jonas BRAASCH[(2)]

[(1)]Arizona State University, U.S.A., Torben.Pastore@asu.edu
[(2)]Rensselaer Polytechnic Institute, U.S.A., braasj@rpi.edu

Abstract

In a series of reports, [1, 2, 3] presented behavioral data for long-duration pairs of noise stimuli presented over headphones with equal but opposite interaural time differences (ITDs). One stimulus led the other by 1-5 ms. Results showed that listeners localized based primarily on the leading noise, even when the lagging noise was as much as 6-8 dB greater in amplitude, thereby demonstrating the precedence effect. Modeling in a recent report (under review) demonstrated how various forms of onset dominance in the extraction of interaural time differences could help to account for these data. However, interaural level differences (ILDs) created by the physical interference of the leading and lagging stimuli appears to have influenced the results for different listeners to varying degrees. The way in which these ITDs and ILDs may be integrated across frequencies and with each other will be explored in new modeling efforts.

Keywords: binaural localization, precedence effect, modeling

1 INTRODUCTION

In the built environment, reflective surfaces such as walls, floors and ceilings introduce early reflections with conflicting spatial cues that might reasonably be expected to make auditory localization impossible. Instead, humans routinely demonstrate a remarkable ability to localize sounds using the spatial cues presented by the first-arriving wavefronts coming directly from the sound source. This ability is called the precedence effect (PE). Study of the PE offers a relatively simple perceptual outcome to a remarkably complex auditory process – the degree to which localization is dominated by the location of the sound source, called *localization dominance*. This outcome facilitates the combined use of psychoacoustical testing and computational modeling to gain insight into temporal aspects of auditory spatial processing. The PE is often studied over headphones by presenting a simulated 'direct sound' (the *lead*) with one combination of ITD/ILD. Then, after a short delay of approximately 1–10 ms, a copy of the lead (the *lag*) is presented with a different combination of ITD/ILD. Listeners can then indicate their perceived lateral position of the sound stimulus using by manipulating the ITD or ILD of another sound, the *acoustic pointer*, to match the lateral position of the sound stimulus.

One aspect of auditory localization that is still relatively poorly understood is how interaural differences of time (ITDs) and level (ILDs) are combined for complex sound stimuli such as those that elicit the PE. In this paper, we consider some strategies for modeling how ITD and ILD might be integrated using physiologically-inspired computational modeling of behavioral PE data using click stimuli collected in our lab. The organizing approach of the present model is to ask what aspects of the stimulus cues would be most salient to an auditory system that seeks to minimize uncertainty in its estimation of an unknown sound source. That is, given the enormous amount of information streaming into the auditory system, what simple mechanisms might help the brain reduce this information to what is essential to successful localization of a target sound? The primary innovation of the model is a mechanism that selects only those time slices which have the most salient cues. Saliency is determined by the relative height of the cross-correlation pattern in one time-frequency bin as compared to others within recent time. By focusing on cues with a saliency above a certain threshold, the information in the signal that is used for lateralization of the auditory event is greatly reduced.

2 Model Structure

Figure 1 shows a schematic diagram of the basic precedence effect model structure in terms of putative peripheral and midbrain sites of auditory processing. The input to the model is the stimulus pressure waveforms at the left and right ears. Central to the approach of this model is the assumption that stimulus binaural cues in each critical band will contribute to the ultimate decision variable to a degree that is roughly proportional to their spectral energy relative to the cues contained in other frequency bands. That is, the spectral regions of the stimulus that have the most energy are likely to be the most salient, and therefore the binaural cues contained in these regions are likely dominate the decision variable. For this reason, the binaural cross-correlation is not normalized by its energy, so that binaural coherence AND the energy in that band serve together to determine the relative saliency of the ITD estimation in that frequency band. Figure 2 shows a schematic representation of the signal processing stages for ITD estimation. Detailed description of key individual components follows.

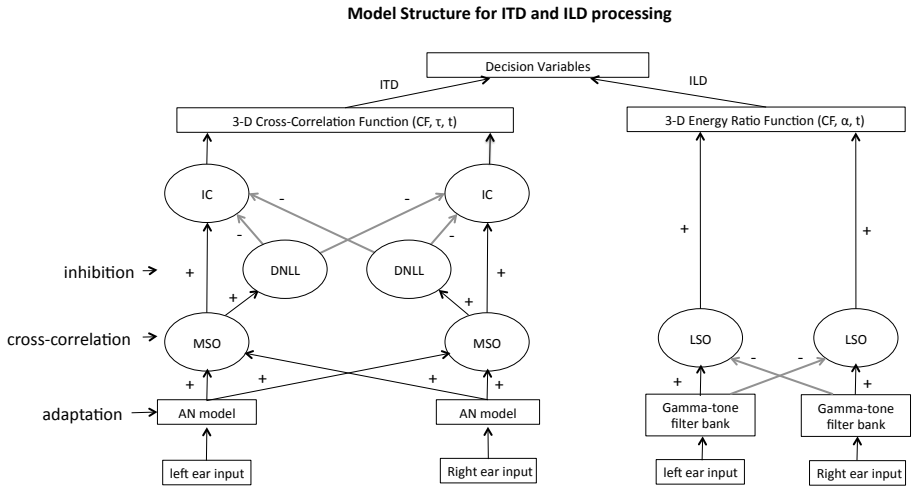

Figure 1. A schematic description of the precedence effect model. Acoustic input at the two ears is transformed into a neural rate function by the AN model. Coincidence detection is performed at the MSO, resulting in excitatory (ipsilateral) and inhibitory (ispsilateral and contralateral) projections to the IC, where they are combined to produce a binaural display as a function of CF, interaural delay, and running time. For ILDs, the acoustic signal is passed through a gammatone filterbank. Long- and short-term ILDs are computed for each filter band in the modeled LSO. ITD and ILD estimates are combined across CFs to produce respective ITD and ILD estimates. These are in turn combined to produce a prediction of the perceived lateral extent of the stimulus in the form of a decision variable.

2.1 ITD Estimation

The AN model of Carney and colleagues simulates the adaptation occurs at the IHC-AN synapse. This adaptation process operates in a very non-linear manner, where firing rate decreases over time in response to a constant stimulus. The model of [4] includes integrated models of the spectral filtering of the middle ear and a non-linear bandpass filter to model the basilar membrane . The bandwidths of the modeled filters vary with amplitude fluctuations to include the compressive non-linearity in BM mechanisms in the output. Inner hair cell transduction is modeled as a non-linear compressive function that saturates at high stimulus intensities, resulting in decreased synchrony to the stimulus input. The output of the Carney model is the instantaneous, time-varying probability of the occurrence of a spike in the modeled AN fiber at the center

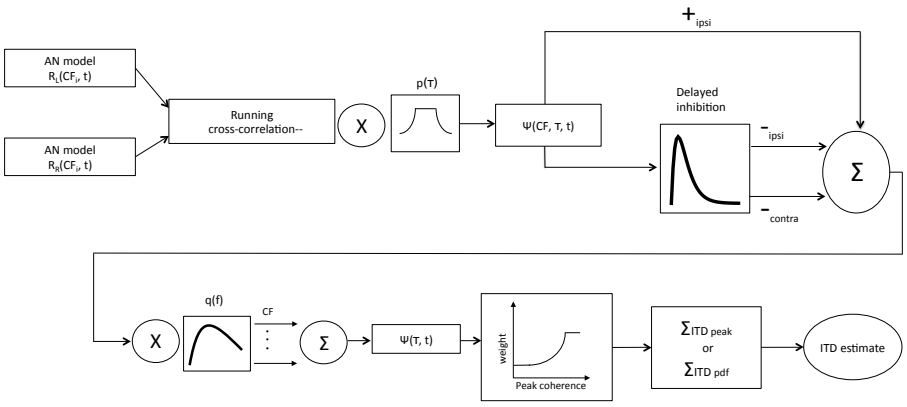

Figure 2. Schematic diagram of signal processing for ITD estimation. The running cross-correlation of the AN model outputs is calculated using a 4 ms time window, advanced by 250 μs for each estimate. The result is then multiplied by a variation of the $p(\tau)$ of [5] . Direct excitatory output is combined with cumulative, delayed ipsilateral and contralateral inhibitory outputs. The resulting cross-correlation function is multiplied by one of several frequency-weighting functions, $q(f)$, and then summed across CFs. The time slices with a peak binaural coherence above the pre-set coherence threshold are selected. Their corresponding cross-correlation functions are then weighted by their relative binaural coherence following a thresholded transfer function. A weighted average is then calculated to produce an ITD estimate in the form of a pdf (the combined cross-correlation function) or a single number estimate (the weighted average of the selected cross-correlation peaks).

frequency of interest. The covariation of stimulus intensity and adaptation was one of the primary reasons for the modeling decision not to normalize the cross-correlation function and to select those time-frequency bins with the greatest energy-coherence.

Figure 3 (left panel) shows the effect of peripheral processing on the internal representation of ITD at the level of MSO. What can be seen is that even though the leading and lagging clicks are of the same amplitude and spectral content, the response to the lag is reduced compared to the lead. If the auditory system is sensitive to the relative excitation at ITD-sensitive neurons, then it might be expected that perceived lateral position will be dominated by the leading click, as is indeed the case for behavioral measures of this stimulus.

The excitatory MSO output explains behavior for many PE conditions, but often fails when the intensity of the lag exceeds that of the lead. This suggests that the PE may not be explained by peripheral processing alone. While MSO has direct excitatory projections to IC, it also projects to the dorsal nucleus of the lateral lemniscus (DNLL). The DNLL responds to excitation from MSO by sending inhibitory projections to both ipsilateral and contralateral IC. A simplified interaction between excitation and inhibition at IC is described as follows. At any time t_n, at any internal delay τ, the ipsilateral IC receives direct excitation from ipsilateral MSO and combined, cumulative inhibition from both ipsilateral and contralateral DNLL. At any time t_n, at any internal delay τ, the activities of both ipsilateral and contralateral DNLL are proportional to the ipsilateral MSO excitation from time t_0 through t_{n-1} weighted by a decaying α function.

The resulting excitatory IC binaural display, E' is $E'(t_n,\tau) = E(t_n,\tau) - I(t_n,\tau)$, where cumulative inhibition $I(t_n,\tau)$ is defined as $I(t_n,\tau) = I_{ipsi}(t_n,\tau) + I_{contra}(t_n,\tau)$. The ipsilateral and contralateral components of inhibition are defined as $I_{ipsi}(t_n,\tau) = E(t_0,\tau)\alpha(t_0) + E(t_1,\tau)\alpha(t_1) \ldots + E(t_{n-1},\tau)\alpha(t_{n-1})$, $I_{contra}(t_n,\tau) = I_{ipsi}(t_n,-\tau)$. The delayed, long-lasting inhibition α, where τ_{peak} is the time delay at which inhibition reaches its peak strength, is modulated by an alpha function of the form

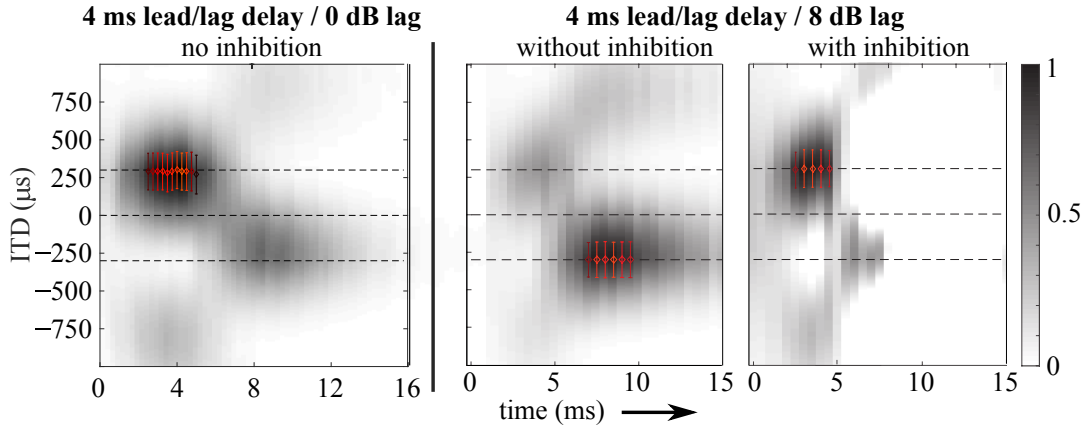

Figure 3. **Left panel:** The running cross-correlation is shown, with internal delays indicated along the ordinate. The height of the combined cross-correlation is indicated by the shade of gray, with black as the highest relative energy-coherence. Time slices with energy-coherence greater than the threshold value (0.7 in this case) are highlighted in red. The center of each red line indicates the peak of the cross-correlation in that time-slice, and the errorbars show the relative width of the function. Positive internal delays are correlated with the lead side. **Right panels:** The same stimulus as the panel, but the lagging click is 8-dB greater amplitude than the lead. On the left is the purely excitatory output of the modeled MSO cross-correlation function. The ITD at the lag position dominates. On the right, the resulting internal display of ITDs at the level of modeled IC, which includes the combined effects of excitation with ipsi- and contralateral inhibition via the modeled DNLL.

$$\alpha(t-t_0) = \alpha_{max}\frac{t-t_0}{\tau_{peak}} \times \exp\left[1-\frac{t-t_0}{\tau_{peak}}\right], \tag{1}$$

This mechanism is only an abstraction at the simplest level of what this inhibition would be at the physiological level and no real effort was made to tune it. This was purposeful. The aim of including this mechanism in the present model was to determine whether hair-cell adaptation *in itself* is enough to explain precedence, especially when the intensity of the lag exceeded that of the lead. Listener responses to click stimuli where lag intensity was increased still showed localization dominance and hair-cell adaptation did not give this same result, suggesting that this further level of inhibition may be necessary.

The right two panels in Fig. 3 show an example of the inhibition generated by excitatory MSO output and the effect it ultimately has on the internal representation of lateral position. At the level of the modeled MSO, the effects of adaptation are not strong enough to favor the lead. Instead, the lag dominates the internal representation of ITD. Introducing delayed, long-lasting inhibition at the level of IC reduces the excitation at the ITD of the lagging click, so that the leading click is likely to dominate perceived lateral position, as was most often the case in [1].

2.1.1 Non-Normalized Cross-Correlation – Energy-Coherence Weighting

The same experiment of [7] that established the spectral dominance hypothesis, showed that increasing the relative level of a stimulus band increases the effect the ITD in that spectral region has on the overall lateralization of a broadband stimulus. [8] weighted the combined ITD and ILD cues by the relative energy in each time/frequency bin in the running analysis, so that cues contribute to the decision variable proportional to the energy in their frequency region. Because the basic philosophy of the current model is that those

cues that are most salient will dominate lateralization, it was decided to emulate the approach of [8]. This was done by simply not energy normalizing the cross-correlation function. Therefore, the height of a given cross-correlation peak in this model is the result not only of how similar the two signals are at that internal delay, but also of the relative energy in that band. So that these values could be compared and used for weighting of the ITD estimates across frequency and time, all cross-correlation functions were normalized by the highest peak value across all frequency bands and all time slices.

ITDs were integrated across frequency based on the 'spectral dominance' region [7] between approximately 600-800 Hz as modeled by [6] (the $q(f)$ function) was used to weight ITDs across frequency. After the $q(f)$ frequency weighting was applied to ITDs, all the cross-correlation functions were simply added across frequency in each time slice to create a compound running cross-correlation. To determine which parts of the integrated cross-correlation function would be most 'salient' to the determination of the ITD estimation, an energy-coherence threshold was set, so that only the cross correlations from time slices that had a peak value greater than the energy-coherence threshold were considered. The cross-correlation functions were then weighted by their peak energy-coherence values, and a weighted average of the above-threshold cross-correlation functions was calculated.

2.2 ILD Estimation

Figure 4 shows a schematic representation of the signal processing used to create a modeled internal representation of ILDs. To simulate the compressive function between ILD and perceived laterality over headphones documented by [9], an error function was used, tuned so that the mean value of ILD (presented with an ITD of 0) that elicited the same perceived lateral position as an ITD of 300μs (presented with 0 ILD) measured across 16 listeners in [1] (5.2 dB) was mapped to 300μs in the model.

$$ILD_{remap} = \frac{2}{\sqrt{\pi}} \int_0^{ILD/19.1} e^{-t^2} dt \tag{2}$$

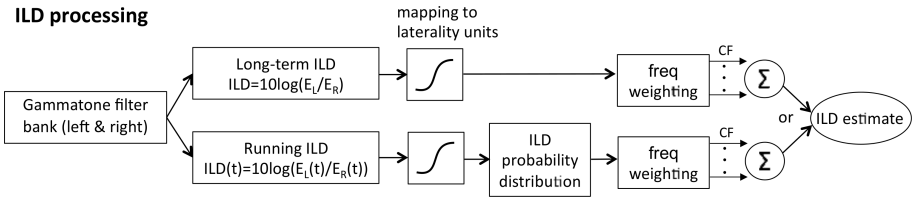

Figure 4. Schematic diagram of signal processing for the estimation of ILDs. The stimulus is filtered to the same CFs as for the AN model using a gammatone filterbank. For the long-term estimate, the decibel ratio of the entire left and right stimulus powers is computed. For running estimation, the signal is segmented by an 8 ms time window that is advanced in 50-μs increments, and the decibel power ratio of left and right filter outputs is computed. The resulting long- and short-term estimates are mapped to the same normalized units of laterality used to report the psychophysical data using a compressive error function. A PDF is calculated from the running ILD estimates.

$$\text{MED}_{cf} = \frac{max(E_{L_{cf}}, E_{R_{cf}})}{\sum_{cf=1}^{N} max(E_{L_{cf}}, E_{R_{cf}})}, \tag{3}$$

Short-term ILDs were integrated over an 8-ms duration which was advanced in 250-μs increments (the same as for ITDs). This integration window, twice as long as used for calculation of ITDs, was used to provide a smoothed function of ILD across time. ILDs were then combined into a probability density function of the short-term, running ILD calculations for each modeled auditory filter. This was in turn normalized to

have an overall probability of 1. Thus, probability density functions (pdf) for the ITD calculation could be combined and compared with ILD estimations not only in terms of their location, but also with regard to their dispersion. An example of the resulting ILD pdfs can be seen in the left panel of Fig. 5.

It is not clear how ILDs are integrated across frequency by the auditory system. Results of [1] suggest that weighting of ILDs across frequency may differ for different users and in response to different stimuli. For this reason, ILDs were weighted based on the relative distribution of spectral energy in the stimulus. The idea was to weight ILDs by the maximum spectral energy density (MED), in each band, relative to that of the overall stimulus, where $E_{L_{cf}}$ is the energy in the left channel at a specified center frequency. An example of an ILD pdf, integrated across frequency weighted by MED, is shown in the right panel of Fig. 5 with the purple pdf.

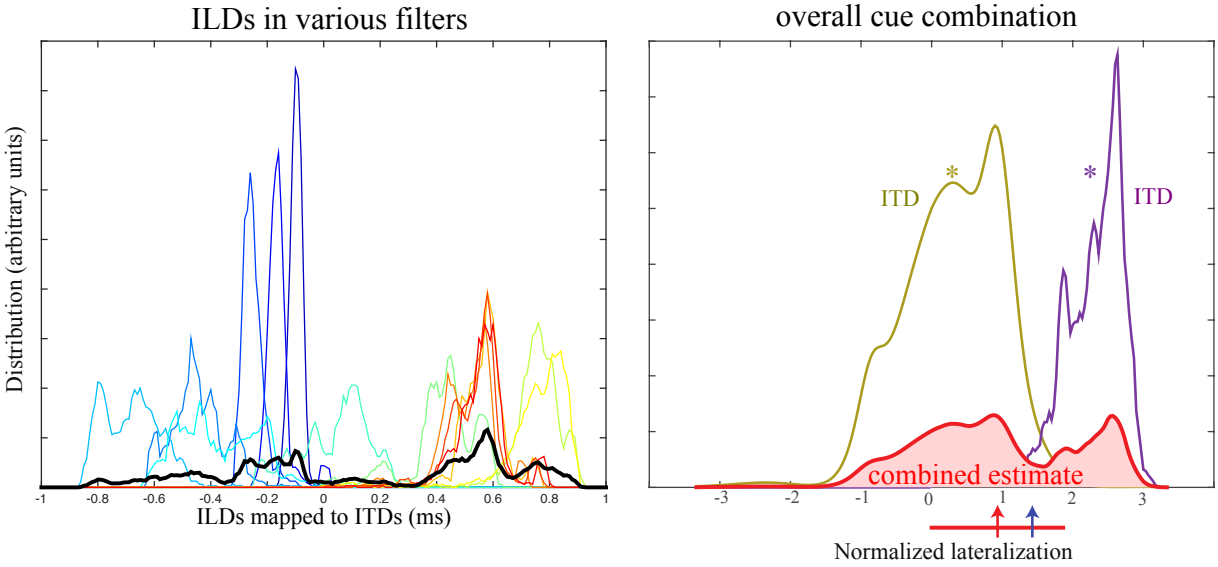

Figure 5. **left panel:** A typical distribution of ILDs over stimulus at different center frequencies. This stimulus was 1 ms ISI, lag level 0 dB. These distributions are then averaged to create a single distribution according to one of the approaches explained in the text. The simple unweighted average is shown by the thick black line in this figure. **right panel:** An example of the pdf internal representation of ITD (tan), ILD (purple), and lateral position (red) estimates derived from the running estimations. The normalized internal representation of laterality is indicated along the abscissa, with positive values to the lead side and negative values to the lag side. The estimated probability of each lateral position is indicated by the ordinate. The final, combined estimate of laterality predicts considerable variability in listener response as the distribution is bimodal – such was the case in the behavioral data for this condition, measured in [1] and [3].

2.3 Combination of ITD and ILD Estimates to Form the Decision Variable

The running estimations of laterality (e.g., Fig. 3, are then transformed (in this case with no temporal weighting) into pdf-based internal representations for ITD and ILD, shown in Figure 5 (right panel) for the 1-ms lead/lag delay with equal level lead and lag (0 dB, the same stimulus condition as in Fig. 3). The final decision variable, shown as a red-filled pdf in the right panel, is the result of the weighted linear addition of the ITD and ILD pdfs, and is smoothed with an 8-point (corresponding to 80 μs) equally-weighted window. This smoothing value was chosen based on the variability found in a behavioral reference condition (see [1]). Tan indicates the ITD estimate, with the tan asterisk denoting the centroid of the ITD distribution. Purple

indicates the distribution of ILDs over the stimulus duration, with the corresponding purple asterisk indicating the centroid of the ILD distribution. The result of weighted linear addition of the ITD and ILD centroids is indicated under the abscissa by the blue arrow. The pdf resulting from the combination of the ITD and ILD pdfs is shown in red, with the centroid and weighted standard deviation indicated by the red arrow and horizontal line under the abscissa. The distribution predicts considerable that there could be considerable variability for this condition, which is what was reported in [3] for this condition.

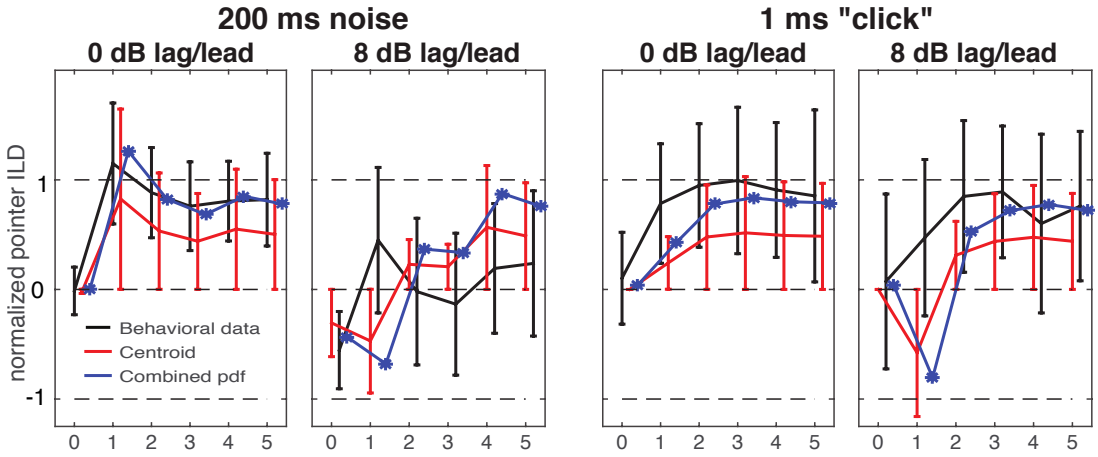

Figure 6. Model outputs for lead/lag pairs of 200-ms noise bursts with 20-ms $\cos^2$ onsets and 1-ms rectangular windowed clicks. Black lines are the mean and standard deviation of listener responses. Model predictions based on the centroid and weighted standard deviation of the model outputs are plotted in red, and referred to as 'centroid estimates.' Blue lines show the additive combination of modeled ITD and ILD distributions, and are referred to as 'combined pdf estimates.' Cyan lines show the combination of the peak ITD and long-term-averaged ILD, and are referred to as 'single-value estimates.'

Figure 6 shows comparisons between the behavioral data reported in [1, 3] and model predictions. The behavioral data are plotted in black with error bars showing the standard deviation of listener responses. The centroid and weighted standard deviation of the overall combined model pdf (the red arrow and horizontal error bar in the right panel of Fig. 5) is plotted in red. The additive combination of the centroids of the ITD and ILD pdfs (the blue arrow in the right panel of Fig. 5) is plotted in blue.

The left two panels of Fig. 6 show behavioral data and model simulation results for a 200-ms duration noise burst, 20-ms $\cos^2$ onset and offset control condition presented in [1, 3]. For presentations at 0-dB lag level, the centroid and single-valued estimates both show the shape of the data very closely . However, the centroid of the overall laterality pdf estimate (red) consistently under-estimates the perceived laterality reported by listeners. Once lag level is increased to 8-dB greater than the lead, both model estimates fail to predict localization dominance for the 1-ms ISI presentations. For greater ISIs, however, predictions based on the centroid and combined pdf predict perceived laterality that is further towards the lead than indicated by the listeners. The variability predicted by the Centroid and weighted standard deviation of the combined pdf estimate of laterality (red error bars) is reasonably close to that reported by listeners in the behavioral data. The right two panels of Fig. 6 show model predictions for the 1-ms rectangular click stimuli. For the most part, the same observations comparing model predictions to the behavioral outcomes for the long-duration noise burst stimuli can also be made for the click stimuli.

3 Summary

This modeling made use of the auditory nerve model of [4] to include effects of peripheral processing on the precedence effect. The hypothesis that the relative energy of both ILD and ITD cues, as well as the binaural coherence of ITD cues (see also [10]) could affect the relative saliency of spatial cues and thus how they are integrated across frequency and with each other was also explored. Finally, the idea of modeling binaural difference cues and perceived laterality as probability density functions, instead of single-valued estimates, was presented. Future efforts will consider how the variability of ITD and ILD estimates may contribute to the temporal weighting of short-term estimates of each of these binaural parameters. In order to better consider the implications and validity of these modeling approaches, it will be helpful to conduct new behavioral experiments where listeners are able to indicate their perceived lateralization not only as a single location, but as a range of locations (e.g., for sound images that are spatially broad or even separated into two locations, for example when the echo threshold is surpassed).

4 Acknowledgements

Portions of this paper were included in Dr. Pastore's 2016 dissertation, "Some effects of the saliency of the lagging stimulus on localization dominance for temporally-overlapping, long-duration noise stimuli," Rensselaer Polytechnic Institute, Troy, NY, 2016.

REFERENCES

[1] Pastore, M.T. and Braasch, J.; 'The precedence effect with increased lag level', The Journal of the Acoustical Society of America, Vol 138 (4), 2015, pp 2079–2089.

[2] Pastore, M.T.,Trahiotis, C., Braasch, J.; The import of within-listener variability to understanding the precedence effect, The Journal of the Acoustical Society of America, Vol 139 (3), 2016, pp 1235-1240.

[3] Pastore, M.T., Braasch, J.; The impact of peripheral mechanisms on the precedence effect, The Journal of the Acoustical Society of America, *under review*.

[4] Zilany, M. S. A., Bruce, I. C., and Carney, L. H.; Updated parameters and expanded simulation options for a model of the auditory periphery, The Journal of the Acoustical Society of America, Vol 135 (3), 2016, pp 283–286.

[5] Colburn, H. S.; Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise, The Journal of the Acoustical Society of America, Vol 61 (3), 1977, pp 525–533.

[6] Stern, R. M., Zeiberg, A. S., and Trahiotis, C.; Lateralization of complex binaural stimuli: A weighted-image model, The Journal of the Acoustical Society of America, Vol 84, 1988, pp 156–165.

[7] Raatgever, J.; On the binaural processing of stimuli with different interaural phase relations, Ph.D. thesis, Delft University of Technology, 1980.

[8] Braasch, J.; A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process, The Journal of the Acoustical Society of America, Vol 134 (1), 2013, pp 420–435.

[9] Yost, W. A.; Lateral position of sinusolds presented with interaural intensive and temporal differences, The Journal of the Acoustical Society of America, Vol 70, 1981, pp 397–409.

[10] Faller, C. and Merimaa, J.; Source localization in complex listening situations: Selection of binaural cues based on interaural coherence, The Journal of the Acoustical Society of America, Vol 116, 2004, pp 3075-3089.