Binaural cues for detection of signals in noise: Experiments and models

Laurel H. Carney

Departments of Biomedical Engineering and Neurobiology and Anatomy, University of Rochester, Rochester, NY, USA

One of the most important functions of the binaural auditory system is detection in noise. Binaural cues provide a significant advantage over the cues available for monaural or diotic detection. However, the details of which cues, or combinations of cues, are used by listeners for binaural detection is still not well understood. Experiments and models that have used reproducible (frozen) noise maskers that allow close examination and manipulation of binaural cues for detection, focussing on recent work on this topic from our group [Davidson *et al*., J. Acoust. Soc. Am. **126**, 1889-1905 (2009); Davidson *et al*., J. Acoust. Soc. Am. **126**, 1906-1925 (2009)], are briefly reviewed here. Experimental results show that a nonlinear combination of fine-structure and envelope cues is required to explain the performance of listeners. Most models for binaural detection depend strongly on either interaural time differences or interaural level differences, or on the variations of these cues over time. These models do not display the same type of interaction of these cues that is observed for human listeners.

INTRODUCTION

Difficulty hearing in noisy situations is generally regarded as the most significant deficit for listeners with hearing loss. A better understanding of how the healthy auditory system copes with complex and noisy environments may lead to improved strategies for preserving or enhancing the most valuable acoustic cues by assistive hearing devices. Studies of detection of simple signals, such as tones, in noise maskers have identified a number of potential cues for detection. Our recent studies have attempted to determine which of these cues are dominant when multiple cues are present, and how these cues may interact. Here we will focus on binaural detection, or detection of a tone that contains binaural differences in the presence of a diotic masker. In particular, we will focus on the case in which a 500-Hz tone is inverted to one ear, and identical narrowband noise maskers are presented to both ears, referred to as the $N_0S\pi$ condition. Subjects with normal hearing have the lowest detection thresholds in this condition, with an advantage over diotic detection of approximately 10-25 dB, depending on bandwidth and varying considerably from subject to subject.

One strategy for studying detection in noise is to use reproducible noise waveforms as maskers, and to take advantage of the significant differences in the amount of masking that occur across different masker waveforms. Several studies have applied this approach to binaural detection (Gilkey *et al*., 1985; Siegel and Colburn, 1989; Isabelle and Colburn, 1991; Isabelle, 1995; Isabelle and Colburn, 2004; Evilsizer

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

et al., 2002; Davidson *et al*. 2006, 2009a,b). The use of reproducible noise maskers allows the detailed study of the relationship between various cues and the detection performance of listeners. This data challenges models for detection to not only predict average thresholds, but also to predict thresholds on a waveform by waveform basis.

Analysis of detection results for listeners in the $N_0S\pi$ condition with reproducible maskers has shown that several of the classical models for binaural detection, such as normalized or un-normalized cross-correlation models (Osman, 1971) and equalization cancellation models (Durlach, 1963) predict results that are much more strongly correlated to energy differences across masker waveforms than are the listeners' results (Isabelle, 1995; Isabelle and Colburn, 1991, 2004). Variation in binaural cues, which is altered by addition of the target tone to the identical noise waveforms presented to the two ears, provides another potential cue for binaural detection; however, the standard deviation of interaural time and level differences (ITD, ILD) are not able to account for a substantial portion of the variance in performance across waveforms (Isabelle, 1995; Isabelle and Colburn, 2004). Various combinations of ITD and ILDs have also been explored, with the combination of cues occurring either before or after averaging across time during the stimulus (Isabelle and Colburn, 2004; Goupell and Hartmann, 2007).

Another approach to determining which cues are dominant for binaural hearing is to manipulate the cues. For example, ITD and ILD cues can be manipulated individually and/or put in opposition to each other (e.g. van de Par and Kohlrausch, 1997, 1998; Zeng *et al*., 2004). The approach used in the studies reviewed here was to create sets of waveforms for which envelope and fine-structure cues were independently chosen, such that their influence on subjects' results could be determined by comparing results across the sets of waveforms. Envelope and fine structures for narrowband noise waveforms were separated using the Hilbert transform, and recombined to form chimeras (Smith *et al*., 2002). In this way, two sets of waveforms that shared common envelopes but had different fine structures were created, as well as two sets of waveforms that had common fine structure but different envelopes. Energy was equalized across the initial masker waveforms to reduce the influence of energy variation across maskers on the results. It was hypothesized that if ITD cues dominated detection, the detailed experimental results across the two sets of reproducible waveforms that shared fine structures should be strongly correlated; whereas if ILD cues dominated detection, results for the two sets of waveforms that shared envelopes should be strongly correlated. The results of these experiments provide constraints against which a number of models of binaural detection were tested.

METHODS

Experimental Studies

Figure 1 illustrates the methods used to create the chimera stimuli by first separating the fine structure and envelope (Fig. 1A, left), and then re-combining them by taking the product of the envelopes and fine structure extracted from different waveforms

(Fig. 1A, right). For details, including procedures used to remove waveforms that had significant spectral splatter associated with the recombination of envelope and fine structure, see Davidson *et al*. (2009a).

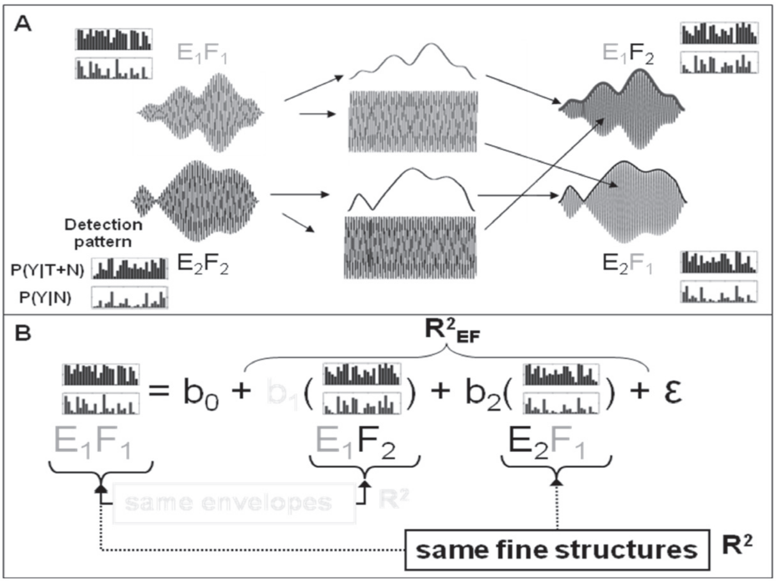

Fig. 1: **A**. Schematic illustration of the procedure for constructing chimeric stimuli. Envelopes (E) and fine structures (F) were separated from the E_1F_1 and E_2F_2 stimulus sets using the Hilbert transform, and then they were exchanged and recombined to create chimeric stimulus sets E_1F_2 and E_2F_1. Detection patterns are shown to remind the reader that each stimulus waveform illustrated is a single member of an ensemble of waveforms. **B**. Illustration of the multiple-regression procedure used to analyze detection results. Chimeric detection patterns sharing envelopes (E_1 in the example above) and sharing fine structure (F_1 above) were used to predict the detection pattern residuals for the baseline stimulus set (E_1F_1 above). The b coefficients represent the slopes of the regression lines used in the multiple regression statistical model; these indicate how strongly the subject weighted the information associated that cue. The b_0 coefficient is always equal to zero because variability linearly associated with the baseline stimulus set not in the model (E_2F_2 above) was removed. The ε term represents error variance. R^2 values were computed for envelope (gray), fine structure (black), and a linear combination of envelope and fine structure (R^2_{EF}). If envelope completely dominated the detection process, it was expected that the E_1F_1 and E_1F_2 detection patterns would be the same and the $R^2_{envelope} = 1$. If fine structure completely dominated the detection process, it was expected that the E_1F_1 and E_2F_1 detection patterns would be the same and the $R^2_{fine\ structure} = 1$. (From Davidson *et al*., 2009a.)

Subjects were extensively trained, first in two-interval tasks with random-noise maskers, then in single-interval tasks with random-noise maskers. After stable performance in the single-interval task was established for tone levels near threshold ($d' \approx 1$), subjects were tested with the reproducible-noise maskers. Values of d′ and bias were monitored; if listener's performance differed significantly from $d'=1$, the tone level was adjusted and testing was re-started. If results indicated bias towards responses of either "Yes" (i.e. "tone present") or "No", subjects were advised accordingly. The experimental procedures were adapted from Davidson *et al*. (2006), Evilsizer *et al*. (2002), and Gilkey *et al*. (1985). The results of the experiments were detection patterns, or hit and false-alarm rates, for each reproducible waveform in the two original (or base line) sets of narrowband stimuli, as well as for the two sets of chimera waveforms (shown in Fig. 1A). Detection patterns were constructed for the probability of "Yes" (Y) responses for tone-plus-noise (T+N) stimuli [P(Y|T+N), i.e. hits], or for noise-alone (N) stimuli [P(Y|N), i.e. false alarms]. The first probability in each P(Y|N) detection pattern is the probability of a Y response for N waveform #1 in that stimulus set, and the first probability in each P(Y|T+N) detection pattern is the probability of a Y response for the T+N stimulus created with N waveform #1 in each set. The second probability in each detection pattern is for N or T+N stimuli created with N waveform #2 in each set, etc. Detection patterns were measured for each subject and for each of the 4 sets of stimuli (2 baseline sets and 2 chimeric sets).

Six subjects completed the experiment. Training and testing procedures were performed in a double-walled sound attenuating booth (Acoustic Systems, Austin, TX). Stimuli were created using MATLAB (Mathworks, Natick, MA) and a TDT System III (Tucker Davis Technologies, Gainesville, FL) RP2 D/A converter (48,828 Hz sampling rate, 24 bits/sample) over TDH-39 headphones (Telephonics, Corp., Farmington, NY).

Detection patterns were analyzed using Pearson's product-moment correlations (r), where r^2 was used to estimate the proportion of variance explained. Multiple regression analyses were used to determine the relative contributions of envelope and fine-structure cues to the subjects' results. These relative contributions were determined by the success of predicting the detection pattern for one set of stimuli based on the detection pattern for another set of stimuli that shared either envelope or fine structure (for details, see Davidson *et al*., 2009a).

Modeling Studies

The experimental results described above, along with related results, were used to test several models for binaural detection. Some of the models were based on decision variables computed directly from the stimulus waveforms, others were based on signal-processing style models for physiological processing. All of the models tested have previously been shown to have some success in predicting detection thresholds in random noise and/or in a number of masking conditions. However, these models had not previously been tested for stimuli with equal energy, and the signal-processing style models had not previously been tested with reproducible noise results.

The experimental results used to test the models included Isabelle's (1995) results for detection of 500-Hz tones in narrowband maskers (*Study 1*), Evilsizer *et al.*'s (2002) results for both narrowband and wideband maskers (*Study 2*), and Davidson *et al.*'s (2009a) results for narrowband chimeras (*Study 3*).

The decision variables computed directly from the stimuli included energy, the standard deviation of ITD and ILD, and several combinations of ITD and ILD. These combinations included the weighted sum of the standard deviations of ITD and ILD (W_{ST}) and the weighted sum of the average absolute values of ITD and ILD (W_{AV}) (Isabelle, 1995; Isabelle and Colburn, 2004). These two decision variables based on combined cues were referred to as "independent centers" by Goupell and Hartmann (2007), because the two cues were analyzed separately over time before being combined into a single decision variable. In contrast, when ITD and ILD were combined first, and then averaged over time, they were referred to as "auditory image" decision variables (Goupell and Hartmann, 2007). The latter included the average of the standard deviation of the weighted sum of ITD and ILD (X_{ST}) and the average of the weighted sum of the absolute values of ITD and ILD (X_{AV}). Another decision variable explored here that combined ITD and ILD cues was the lateral position (L_P) decision variable, in which a trading ratio is used to combine ITD and ILD (Hafter, 1971).

Two signal-processing style models were also tested with the reproducible-noise results. Marquardt and McAlpine's (2001) four-channel model incorporates two channels that compute cross-correlations between a model auditory-nerve (AN) response from one ear and a signal from the opposite ear that is delayed by $\pi/4$ radians at 500 Hz (i.e. ±250 μsec) and two channels that compute the difference between the model AN response from one ear and the $\pi/4$-delayed response of the opposite ear (Fig. 2). This model was inspired by the finding that when ITD curves for binaural neurons are examined in groups that are tuned to similar frequencies, the best ITDs tend to cluster at delays that correspond to phase shifts of $\pi/4$ for the best frequency (McAlpine *et al.*, 2001). The other signal processing model tested was that of Breebaart *et al.* (2001), which is comprised of a simple model for the auditory periphery, including adaptation loops, and a binaural processor based on excitatory-inhibitory (EI) interactions (Fig. 2).

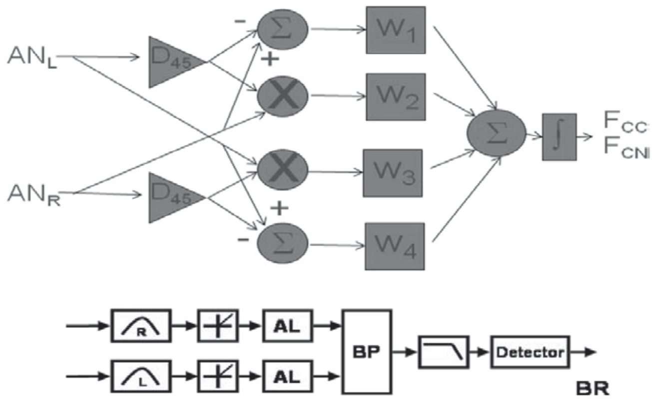

Fig. 2: Block diagrams of the four-channel (FC) and binaural Breebaart (BR) models used to predict N_0S_π detection patterns. D_{45} denotes a delay block corresponding to a phase delay of 45 degrees based on the center frequency of each model auditory-nerve (AN) fiber. AL denotes the adaptation loops as described in Dau *et al.* (1996). BP denotes a binaural processor. (From Davidson *et al.*, 2009b.)

RESULTS

Experimental Results

Figure 3 summarizes the analyses of the experimental data for Davidson *et al.*'s (2009a) experimental results for binaural detection with reproducible chimeric stimuli. The results show the predictions of the residuals of the baseline stimulus set detection patterns based on the listeners' results for the stimulus sets that shared the same envelopes or fine structures. (Residuals were used for these predictions, after partialling out any chance correlations between the stimulus sets, to avoid the possibility of misleading correlations between detection patterns; see Davidson *et al.*, 2009a for details.) Three sets of predictions are shown in the three columns: in the first column, results for the probability of a "Yes" response across all T+N and N waveforms (W) is shown; in the second column, results for the probability of "Yes" for T+N trials are shown, and in the third column, the results for probability of "Yes" responses on N trials are shown.

If the listeners' results depended only on cues associated with the fine structure of the stimuli, and thus dominated by ITD, then the black data points which were predicted based on stimuli that shared the same fine structure would be expected to lie along the diagonal, and the slope of the regression line related to the contribution of the fine structure (b_F, illustrated as b_2 in Fig. 1B) would be close to 1, while the contribution of envelope cues would be minimal (*i.e.* b_E, illustrated as b_1 in Fig. 1B, would be close to 0.) Similarly, if the cues associated with envelope, and thus dominated by ILD, were dominant, then the gray symbols, predicted based on responses to stimuli that shared the same envelopes, should lie along the diagonal line, b_E should be

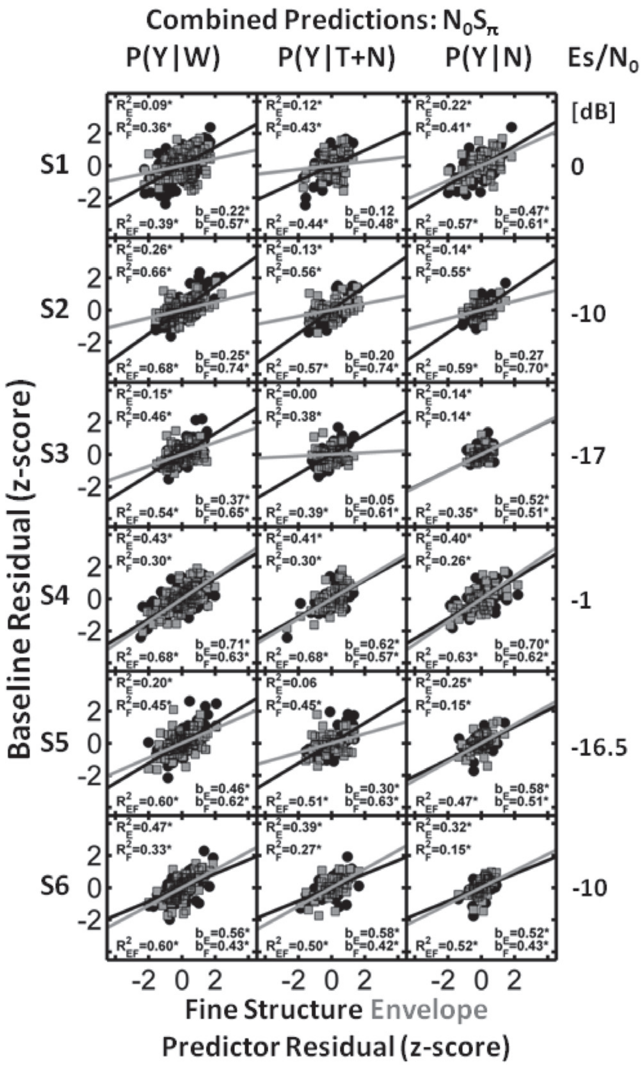

Fig. 3: Predictions for responses to the E_1F_1 and E_2F_2 $N_0S\pi$ stimulus ensembles for 6 subjects (rows) based on envelope(grey squares) or fine structure (black circles). Relative weights for each cue are shown by the b_E and b_F values, with asterisks indicating significant slopes. R^2_E and R^2_F indicate the proportion of predictable variance based on the individual cues (asterisks indicate significant values). R^2_{EF} corresponds to the proportion of predictable variance using a linear combination of both envelope and fine structure. Signal-to-noise ratio (E_S/N_0 *in dB*) is shown to the right of each row. (From Davidson *et al*., 2009a.)

close to 1, and b_F should be close to 0. The results show that, in general, neither fine structure or envelope cues dominated the subjects' responses, but rather that both cues contributed to the results. Furthermore, the proportion of variance predicted by the linear combination of both cues, R^2_{EF} (see Fig. 1B) are relatively low compared to estimates of predictable variance for these data, which range from 0.80 to 0.99 (predictable variance can be estimated based on the reliability of each subject's detection patterns, see Davidson *et al*., 2009a,b). This result suggests that while both cues contribute to the responses, they must interact in a nonlinear manner.

Modeling Results

The detection patterns described above, as well as those from two other studies of binaural detection in reproducible noise, were predicted using the decision variable and signal-processing style models described above. The results are shown in Fig. 4 as the proportion of variance in each subject's detection pattern (i.e. the variability in hit rates across different reproducible stimuli) that was explained by each model. Only hit rates could be studied with the decision variables based on binaural differences, because for N trials the differences between the stimuli to the two ears were identically zero.

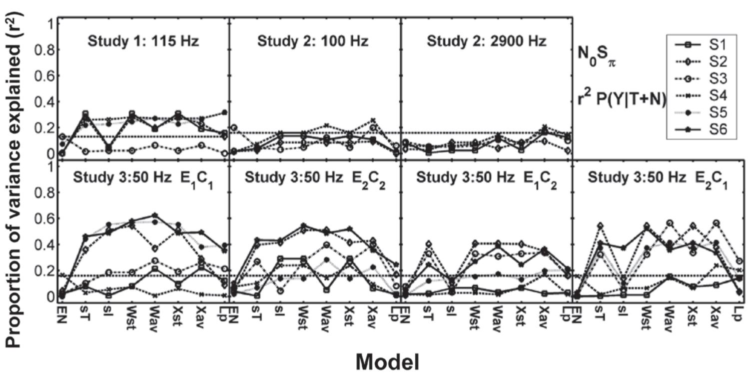

Fig. 4: Proportion of variance explained (r^2) by selected Isabelle (1995), Goupell (2005), and Hafter (1971) decision variables for *z* scores of P(Y|T+N). EN is the RMS energy of the right stimulus waveform; sT is the standard deviation of ITDs; sI is the standard deviation of ILDs; Wst is a linear combination of the standard deviations of ITDs and ILDs; Wav is a linear combination of the average absolute values of ITDs and ILDs; Xst is the standard deviation of a linear combination of ITDs and ILDs; Xav is the average value of a linear combination of ITDs and ILDs; and Lp is a lateral position model relating ITDs and ILDs with a trading ratio. The critical r^2 value for reaching a significant prediction ($p < 0.05$), including a Bonferroni correction for comparison of each data set to 11 models is 0.29 for *Study 1* and 0.30 for *Studies 2* and *3*, as denoted by the horizontal dashed lines. Symbols are connected to facilitate comparisons across models. (From Davidson *et al*., 2009b.)

The proportion of variance in the detection patterns that was explained by ITD- and ILD- related decision variables was relatively low for all subjects in *Studies 1* and *2*, and for 3 of the subjects in *Study 3* (Fig. 4). For the other three subjects, decision variables involving combinations of ILD and ITD provided the best predictions of experimental results, although these predictions were in general not significantly higher than those based on only the standard deviation of ITD

(sT) or ILD (sI). Recall that the combined decision variables included a weighted sum that was fitted to the subject data. The values of the resulting weights are revealing (Fig. 5). These decision variables tended to have weights near 1 or 0, indicating that these models tended to select the better cue for each subject, despite the efforts to devise decision variables that took advantage of both ITD and ILD cues.

The results for predictions of the subject's detection patterns using the two signal-processing-style models described above are shown in Fig. 6, which shows the proportion of variance in the detection patterns that was explained by model predictions. For nearly all cases tested, the predictions were quite poor, only reaching statistical significance in a few cases for the binaural Breebaart model. These models represent attempts to include physiologically motivated mechanisms, such as peripheral processing followed by cross-correlation and/or interactions between excitation and inhibition. These results make it clear that our understanding of the neural mechanisms involved in binaural processing is not yet adequate to fully understand listener responses at the level of detail provided by detection patterns for reproducible noise.

Using model predictions for the reproducible chimera waveforms in *Study 3*, it was possible to analyze the relative dominance of fine-structure versus envelope cues in the model responses. Unlike the human listeners, who apparently make use of both aspects of the available temporal cues (see Fig. 3), the models tended to be strongly dominated by one or the other (see Davidson *et al*., 2009b for details.)

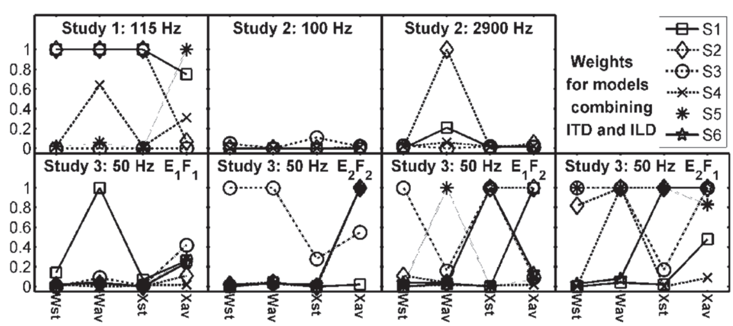

Fig. 5: Model weights for decision variables based on both ITDs and ILDs. A weight approaching 1 indicates reliance on ITD and a weight approaching 0 indicates reliance on ILD. Different subjects are indicated with different symbols connected with lines to facilitate intersubject and cross-model comparisons. (From Davidson *et al*., 2009b.)

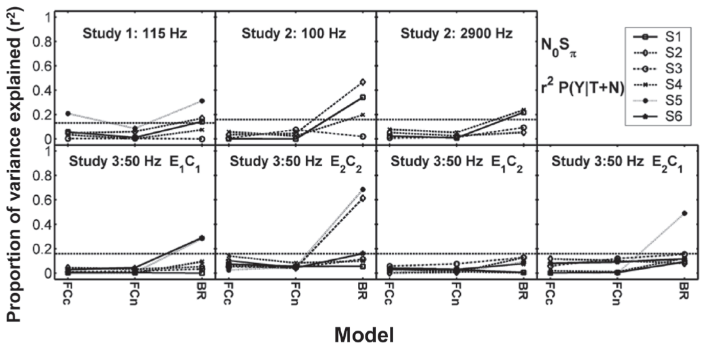

Fig. 6: Proportion of variance explained (r^2) by N_0S_π model predictions for z scores of A) P(Y|T+N) and B) P(Y|N). Model abbreviations are: FCc, four-channel model using an unnormalized cross correlation for simulated peaker channels (i.e. channels with ITD curves that are characterized by a central peak); FCn, four-channel model using a normalized cross correlation for peaker channels; BR, the binaural Breebaart model. Symbols are connected to facilitate comparisons across models. The critical r^2 value for reaching a significant prediction ($p < 0.05$), including a Bonferroni correction for comparison of each data set to 11 models for P(Y|T+N), is 0.29 for *Study 1* and 0.30 for *Studies 2* and *3*; and including a correction for comparison to 2 models for P(Y|N) the critical r^2 value is 0.20 for all three studies, as indicated by the horizontal dashed lines. (From Davidson *et al*., 2009b.)

DISCUSSION

The results of the experimental and modelling studies reviewed here illustrate the type of information that can be gained using reproducible noise studies. Detailed sets of responses for sets of specific waveforms provide new information about the cues that listeners use for this task. The use of chimeric stimuli provided insight into the interaction of those cues; in particular, temporal cues associated with the envelope and fine structure of the stimuli both influenced the listeners' responses, and linear combinations of these cues were not successful in predicting the response patterns. These data also provide a challenge for models of binaural detection. Many existing models focus on one particular cue (e.g. ITD- or ILD-related cues). Even those that intend to take advantage of two cues tend to be dominated by the better of the two cues. These aspects of existing models do not agree with the patterns that were observed in the listeners' responses.

Future models for binaural detection must explore and incorporate the interactions between envelope and fine-structure features of the stimuli. One possible approach is to include more biophysically realistic representations of neural mechanisms, which naturally introduce interactions between energy and timing. For example, models of the auditory periphery with level-dependent tuning introduce interactions between envelope and timing of model AN responses, and neural models for coincidence detectors are strongly influenced by both the number and relative timing of action potentials.

ACKNOWLEDGEMENTS

This work was supported by NIH-NIDCD R01001641. Drs. Sean Davidson, R. H. Gilkey, and H. Steven Colburn collaborated on the experimental and modelling studies reviewed here, which were the basis for Sean Davidson's dissertation at Syracuse University (Davidson, 2007).

REFERENCES

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

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615-3622.

Davidson, S. A. (**2007**). *Detection of Tones in Reproducible Noise: Psychophysical and Computational Studies of Stimulus Features and Processing Mechanisms,* Unpublished PhD Dissertation, Syracuse University (www.bme.rochester.edu/carney.html).

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (**2006**). "Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results," J. Acoust. Soc. Am. **119**, 2258-2275.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (**2009a**). "Diotic and dichotic detection with reproducible chimeric stimuli," J. Acoust. Soc. Am. **126**, 1889-1905.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (**2009b**). "An evaluation of models for diotic and dichotic detection in reproducible noises," J. Acoust. Soc. Am. **126**, 1906-1925.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206-1218.

Evilsizer M. E., Gilkey R. H., Mason C. R., Colburn H. S., and Carney L. H. (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336-345.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207-1219.

Goupell, M. J., and Hartmann, W. M. (**2007**). "Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models," J. Acoust. Soc. Am. **122**, 1029-1045.

Hafter, E. R. (**1971**). "Quantative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116-1122.

Isabelle, S. K., and Colburn, H. S. (**1991**). "Detection of tones in reproducible narrow-band Noise," J. Acoust. Soc. Am. **89**, 352-359.

Isabelle, S. K. (**1995**). *Binaural detection performance using reproducible stimuli,* Ph.D. Dissertation (Boston University).

Isabelle, S. K., and Colburn, H. S. (**2004**). “Binaural detection of tones masked by reproducible noise: Experiment and models,” Report BU-HRC 04-01.

Marquardt, T., and McAlpine, D. (**2001**). “Simulation of binaural unmasking using just four binaural channels,” J. Assoc. Res. Otolaryn. (Abs. 21716)

McAlpine, D., Jiang, D., and Palmer, A. R., (**2001**). “A neural code for low-frequency sound localization in mammals,” Nature Neuroscience **4**, 396-401.

Osman, E. (**1971**). “A correlation model of binaural masking level differences,” J. Acoust. Soc. Am. **50**, 1494-1511.

Siegel, R. A., and Colburn, H. S. (**1989**). “Binaural processing of noisy stimuli: Internal/external noise ratios under diotic and dichotic stimulus conditions,” J. Acoust. Soc. Am. **86**, 2122-2128.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). “Chimaeric sounds reveal dichotomies in auditory perception,” Nature **416**, 87–90.

van de Par, S. and Kohlrausch, A. (**1997**). “A new approach to comparing binaural masking level differences at low and high frequencies,” J. Acoust. Soc. Am. **101**, 1671-1680.

van de Par, S. and Kohlrausch, A. (**1998**). “Diotic and dichotic detection using multiplied-noise maskers,” J. Acoust. Soc. Am. **103**, 2100-2110.

Zeng, F., Nie, K., Liu, S., Stickney, G. Del Rio, E., Kong, Y., and Chen, H. (**2004**). “On the dichotomy in auditory perception between temporal envelope and fine structure cues,” J. Acoust. Soc. Am. **116**, 1351-1354.