

# Trading of intensity and interaural coherence in dichotic pitch stimuli

John F. Culling<sup>a)</sup> and Hannah G. Lewis

*School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom*

(Received 21 July 2009; revised 21 July 2010; accepted 22 July 2010)

When a signal is added to noise in the NoS $\pi$  binaural configuration, a reduction in interaural coherence,  $\rho$ , occurs at the signal frequency and increases in tone intensity decrease  $\rho$ . Corresponding manipulations of  $\rho$  result in the perception of a phantom signal which increases in loudness as  $\rho$  decreases [Culling *et al.* (2001). *J. Acoust. Soc. Am.* **110**, 1020–1029]. In the present study, a narrow sub-band of noise (462–539 Hz) embedded within a broadband (0–3 kHz) diotic noise was manipulated in both intensity and  $\rho$  in a 3-interval, odd-one-out task. In the reference intervals,  $\rho$  was zero and the spectrum was flat. In the target interval, both  $\rho$  and the intensity of the target band were incremented giving opposing effects on loudness. Correct identification of the target interval followed a V-shape as a function of the size of intensity increment. The minimum of this function was often at chance performance, indicating that monaurally and binaurally evoked loudness were fully traded. These results show that reduction in  $\rho$  at a given frequency produces increased loudness at that frequency equivalent to up to 6 dB and consistent with an equalization-cancellation mechanism whose binaural output is strongly weighted compared to monaural excitation. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3478853]

PACS number(s): 43.66.Cb, 43.66.Pn, 43.66.Ba [MAA]

Pages: 1908–1914

## I. INTRODUCTION

Dichotic pitches are typically heard when random noise, applied to both ears, has a reduced interaural coherence at one or more spectral frequencies (Durlach, 1962; Culling *et al.*, 1998a, 1998b). Interaural coherence is defined here as the maximum of the waveform interaural cross-correlation within a given frequency channel.<sup>1</sup> The perceived pitch depends upon the identity of those spectral frequencies in a similar way to the dependence of monaural pitch on the monaural frequency spectrum, including the pitch of the missing fundamental (Bilsen, 1977; Gockel *et al.*, 2009) and “de Boer’s rule” (Culling *et al.*, 1998a). This correspondence implies the existence of a central spectrum that somehow represents reductions in interaural coherence as a function of frequency. Percepts of pitch and timbre may be derived from this central spectrum in a similar way to the corresponding monaural percepts, but limited by the effective frequency range of the binaural system (Hirsh and Burgeat, 1958; Cramer and Huggins, 1958; Culling, 1999). These phenomena show a close correspondence with those of binaural unmasking (Hirsh, 1948; Durlach and Colburn, 1978), suggesting that they are manifestations of the same underlying mechanism (Durlach *et al.*, 1986; Culling 2007).

The salience of dichotic pitches, and thus the dynamic range of this central spectrum, has rarely been addressed. Most studies both of dichotic pitches and of binaural unmasking have been concerned only with detection. This omission is significant, because the identification of complex sounds, such as speech, in background noise requires the recovery and interpretation of a spectral profile rather than a

binary present/absent decision. In order to characterize this spectral profile, we need to know how excitation, and so loudness, grows as a function of changes in interaural coherence. Only a few studies have evaluated the salience of dichotic pitches. These have included measures of their relative ability to evoke recognizable melodies (Akeroyd *et al.*, 2001), measures of discrimination sensitivity between different levels of interaural coherence,  $\rho$ , in a dichotic pitch (Culling *et al.*, 2001) and measures of partial loudness (Gockel *et al.*, 2009).

The melody recognition task of Akeroyd *et al.* provided a hierarchy among three forms of dichotic pitch in comparison with the melody recognition facilitated by pure tones. Pure tones provided the most accurate melody recognition, followed by Huggins pitches and then binaural edge pitches (Klein and Hartmann, 1981). The binaural coherence edge pitch (Hartmann and McMillon, 2001) provided markedly lower levels of recognition, although identification rates were still 70%–80% correct. This hierarchy is consistent with idea that the salience of a dichotic pitch is related to the reduction in  $\rho$  at the spectral frequency concerned. In the case of the binaural coherence edge pitch, there is no spectral peak in incoherence, but rather an edge, and it is assumed that, in order to evoke a tonal percept, a lateral inhibition process must transform this edge into a peak in the central spectrum. The existence of this rank ordering implies some degree of graded perceptual salience.

Culling *et al.* (2001) attempted to provide a more quantitative assay of the salience of dichotic pitches by focusing specifically on the loudness of the perceived pitch rather than on its tonality. In their “fringed” condition, they measured listeners’ ability to discriminate different levels of interaural coherence,  $\rho$ , in a narrow band of noise embedded within an otherwise diotic broadband noise. The principal perceptual

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: cullingj@cf.ac.uk

effect of changing  $\rho$  appeared to be in the loudness of a whistle-like pitch and listeners were explicitly instructed to perform the task on the basis of the loudness of this whistle. Listeners were sensitive to small changes in  $\rho$  close to a reference value of one, but became less sensitive to changes in  $\rho$  as  $\rho$  was reduced toward zero.

Measurements of  $d'$  between different values of  $\rho$  facilitated the development of a cumulative  $d'$  function that was related to the loudness of the percept. This function had a steeply ascending slope close to  $\rho=1$ , which lessened progressively with decreasing  $\rho$  until it reached its maximum at  $\rho=0$ . This result implies that maximum loudness occurred at  $\rho=0$ . The cumulative  $d'$  at this point was over 6 in some listeners, although the average was 4.5 at 500 Hz. While these data indicate that a band with zero interaural coherence is very perceptually salient in some way, the presumed relationship between this scale and either loudness or the dynamic range of the central spectrum remained speculative. In particular, the fact that dichotic pitches are perceived at a different intracranial position from the rest of the noise (Raatgever and Bilsen, 1977; Akeroyd and Summerfield, 2000; Zhang and Hartmann, 2008), leaves open the possibility that listeners may, despite the instructions, have used the laterality of the whistling sound rather than its loudness.

Gockel *et al.* (2009) measured the partial loudness of a complex Huggins pitch (i.e., a complex of several Huggins pitches at harmonic intervals). The partial loudness is the loudness of the pitch component of the stimulus, rather than the loudness of the entire stimulus. They asked listeners to match the loudness of the tonal part of the stimulus with an equivalent harmonic complex presented in quiet. The loudness was then derived from Moore and Glasberg's (2007) loudness model for this equivalent monaural stimulus. They concluded that the partial loudness of their complex Huggins pitch was equivalent to the same set of harmonics presented at 6.6 dB SL. Again, this figure indicates that Huggins pitches are very salient. Moreover, it relates them, for the first time, to well-established monaural measures of perceptual salience and signal power, although the reliance on a subjective matching procedure still leaves open some possibility that listeners might not have matched purely on the basis of loudness.

The present study set out to achieve a similar end, but without the intervention of a loudness model, and without the use of a matching procedure. The experimental paradigm was, again, based on the idea that dichotic pitches possess some intrinsic partial loudness. It was further supposed that there might be some scope for confusion between the monaural spectral profile of the stimulus and the central spectrum evoked by the interaural relationships. If so, listeners' ability to discriminate between stimuli that differ in monaural spectral profile might be detrimentally affected if the binaurally-evoked central spectrum were to differ in the other direction. That is, if a manipulated sub-band of the stimulus were to increase in intensity at both ears from one presentation interval to another (so increasing in partial loudness), while  $\rho$  for that band also increased from one interval to another (so decreasing in partial loudness), the net result might be a smaller perceptual change than when either cue was manipu-

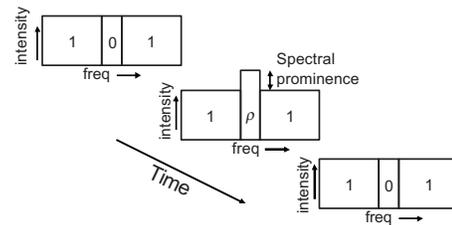


FIG. 1. Schematic illustration of the stimulus and trial structure, with the signal occurring in the second interval. The numbers in each block of the spectrum represent the interaural correlation in that spectral region. Listeners identified the interval which sounded different.

lated independently. The intensity and  $\rho$  were thus anticipated to trade against each other to some extent. Since the design does not rely on instruction in the use of loudness, but on failure to discriminate at all, we can be assured that any such failure can only have occurred if loudness was indeed equalized.

## II. EXPERIMENT

### A. Stimuli and Procedure

Figure 1 illustrates the experimental stimuli and the structure of an individual trial. Each stimulus consisted of a target band of noise flanked by spectrally contiguous flanking bands of diotic noise ranging from 0–3 kHz. The target band was 1 ERB wide (Moore and Glasberg, 1983) and centered at 500 Hz (i.e., 462–539 Hz). Each trial consisted of three 500-ms presentation intervals, separated by 90 ms silences. The target band was manipulated across presentation intervals in  $\rho$  and in intensity. In two non-signal intervals, the target band had the same spectrum level as the flanking bands and  $\rho$  of zero. In the signal interval, the  $\rho$  of the target band was between 0 and 1 and its spectrum level was between 0 and 10 dB higher than that of the flanking noise. This difference in spectrum level is termed the “spectral prominence” of the target band. The signal interval occurred with equal probability as the first, second or third stimulus in a trial. The listeners were instructed to indicate which interval sounded different from the other two. Listeners responded using buttons on a gamepad and received immediate trial-by-trial feedback via a computer monitor.

In order to measure discrimination for a particular combination of  $\rho$  and spectral prominence, listeners were presented with a block of 50 trials with a fixed value of  $\rho$  and a fixed spectral prominence. At the beginning of each block, listeners heard an example of the target band played diotically and in isolation.

Psychometric functions (i.e., performance as a function of spectral prominence) were measured by presenting a series of blocks. Each spectral prominence was tested in a random sequence. For  $\rho \leq 0.8$ , the spectral prominence varied in 0.5-dB steps from 0 to 5 dB. For  $\rho > 0.8$ , spectral prominence varied in 1-dB steps from 0 to 10 dB. Eleven blocks were thus needed to measure one psychometric function. One set of 11 blocks, presented in a random order, made up one 45-min session.

In successive sessions, different psychometric functions for different target-band values of  $\rho$  were measured in a ran-

dom sequence. For purposes of comparison, the same values of  $\rho$  were employed as for the “fringed” condition of Culling *et al.* (2001): these were 0, 0.3, 0.5, 0.65, 0.8, 0.9, 0.95 and 1.0. Listeners completed this set of eight sessions twice, performing no more than one session per day.

Listeners JC and HL (the two authors) were both aware that changes in interaural correlation across the intervals in a trial were likely to affect the loudness and laterality of the target band. HJ and PW were paid for their participation and, aside from having their attention drawn to the target frequency at the start of each block, were told only to identify the interval that differed from the other two. All listeners reported normal hearing at low frequencies (JC had a mild loss at 4 kHz and above).

Stimuli were prepared digitally using MatLab at 44.1 kHz sampling frequency and 16-bit sample depth. All filtering was performed in the frequency domain by zeroing unwanted Fourier components. The interaural correlation of the target band,  $\rho$ , was controlled by mixing two 500-ms bands of Gaussian noises,  $n_1(t)$  and  $n_2(t)$ , according to Eqs. (1) and (2) to create the left- and right-ear noisebands  $l(t)$  and  $r(t)$ ,

$$l(t) = \sqrt{\frac{1+\rho}{2}}n_1(t) + \sqrt{\frac{1-\rho}{2}}n_2(t), \quad (1)$$

$$r(t) = \sqrt{\frac{1+\rho}{2}}n_1(t) - \sqrt{\frac{1-\rho}{2}}n_2(t). \quad (2)$$

Complementary, 500-ms bands of diotic flanking noise (0–462 Hz and 539–3000 Hz) were then added. The completed stimuli were gated by 10-ms, raised-cosine onset/offset ramps before presentation. The experiment was performed in a single-walled IAC sound-attenuating booth using an Edirol UA-20 soundcard, an MTR HPA-2 headphone amplifier and Sennheiser HD650 headphones at an overall sound level of approximately 63 dB(A) when no spectral prominence was added.

## B. Results

### 1. Psychometric functions

Percent discrimination was transformed into  $d'$  using algorithm 2 of Smith (1982) for evaluating  $d'$  in multi-interval, forced-choice tasks. The symbols in Fig. 2 show these  $d'$  values plotted as a function of spectral prominence for each listener at  $\rho=0.65$ , which illustrates the V-shaped data pattern. At zero spectral prominence  $d'$  was between about 1 and 2.5, indicating that a target band with  $\rho=0.65$  was quite easily distinguished from the bands with  $\rho=0$  in the distracter intervals. As spectral prominence was increased, however, discrimination became more difficult, reaching a minimum  $d'$  close to zero at between 1.7 and 2.6 dB spectral prominence, before increasing once again, and reaching more than 2.5 for each listener at a spectral prominence of 5 dB. Here, the peak in spectral power was easily distinguished from the flat spectra of the distracters. The psychometric function was thus V-shaped, suggesting that there is a point of subjective equality (or at least minimal perceptual difference) which occurs at a different level of spectral prominence for each listener. As will be shown below, the

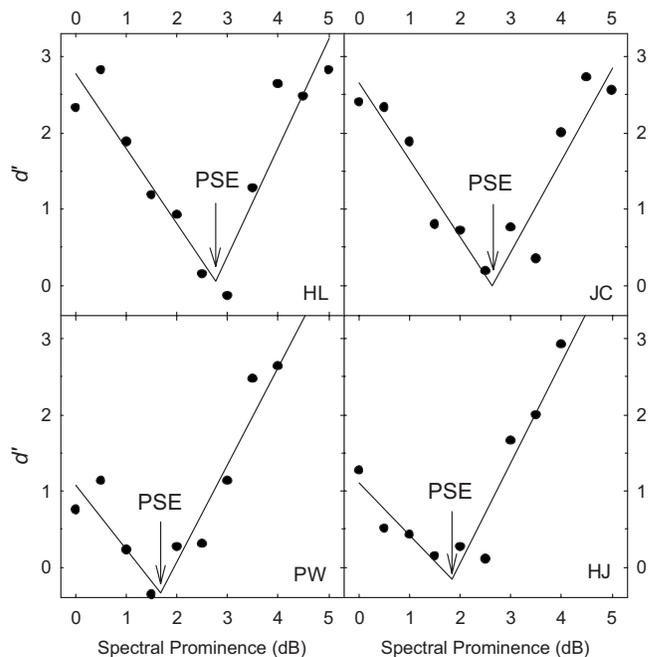


FIG. 2. Example psychometric functions drawn from the raw data (100 trials) of each listener at  $\rho=0.65$ . The symbols are the measured values of  $d'$  and the lines are bilinear fits to this data. The arrows indicate the points of subjective equality derived from the fitted lines.

V-shaped pattern was common to all the data sets and the point of subjective equality (PSE) increased with increasing  $\rho$ .

In order to characterize the V-shape for each listener’s data, bilinear regression fits were applied for each level of  $\rho$  (Fig. 2, lines). Two lines with freely varying slope and y-intercept were fitted such that the sum-squared error between each data point and the uppermost line at the corresponding value of spectral prominence was minimized. The value of spectral prominence at the intercept between these two lines was then taken as the PSE. The raw data are noisy, but the fitted PSE is determined by all the data in the psychometric function, minimizing, in particular, the effect of variability in individual data points close to the PSE. The raw data were also collected at quantized levels of spectral prominence. Consequently, they do not necessarily capture the optimal trading point and the lowest possible value of  $d'$ . The bilinear fits provide an estimate of the minimum  $d'$ .

The PSE was not always accompanied by performance close to chance (see Fig. 4 below). At  $\rho=1$ , in particular, the minimum  $d'$  was always between 0.8 and 1.5 (i.e., 56%–77% correct, where chance=33%), indicating that the stimuli were difficult, but not impossible, to distinguish at the PSE. For  $\rho=0.5$ , on the other hand the minimum in  $d'$  was never greater than 0.2. (38% correct), which was not significantly above chance even with an uncorrected  $\alpha$  (42% needed for  $p<0.05$ ).

### 2. Effect of interaural correlation

The differences in PSE make it difficult to summarize the  $d'$  data, so these were first removed using a normalization procedure. Figure 3 shows the PSEs of individual listeners plotted against the corresponding mean values, averaged

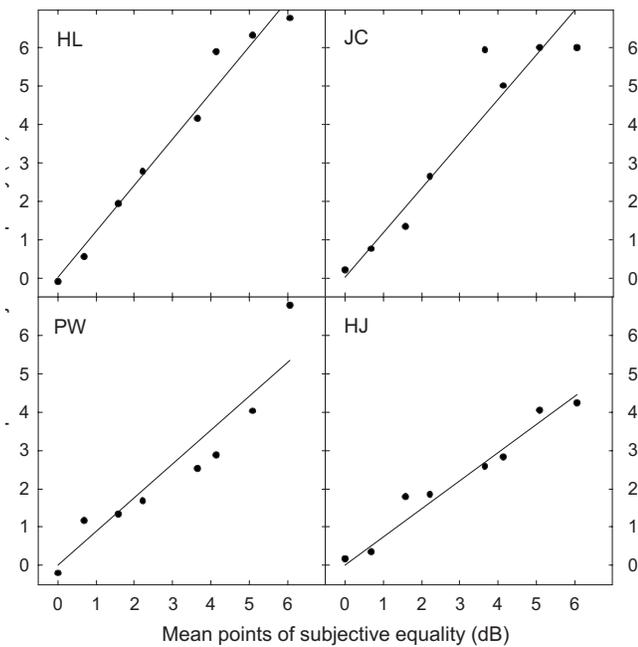


FIG. 3. The symbols show individual points of subjective equality for each listener plotted against the equivalent mean values averaged across listeners. The solid lines are linear regressions to these data, constrained to pass through the origin.

across listeners, accompanied by linear regression lines of variable slope, but fixed at the origin. It can be seen that the differences between listeners can largely be characterized as a simple difference in regression slope. The slopes of these regression lines are 1.21, 1.16, 0.88, and 0.74 for listeners HL, JC, PW and HJ, respectively. The data from each listener could thus be normalized by rescaling the spectral prominence axis using these factors. The normalization of the spectral prominence axis allows the  $d'$  data from different listeners to be overlaid, showing approximately aligned PSEs (Fig. 4). It can be seen that PSE increases with increasing  $\rho$  in a similar way for all listeners. Although this maneuver aligns the V-shaped functions, the measured data points from different listeners do not occur at common points on the abscissa, and so cannot be averaged.

The effect of  $\rho$  on the mean PSE is further illustrated in Fig. 5 (top panel) where the mean fitted PSEs are plotted against  $\rho$ . By extracting the minima from the fitted V-shaped functions, we can obtain an estimate of the minimum discriminability ( $d'$  at PSE) for each value of  $\rho$ . Figure 5 (bottom panel) shows these minimum  $d'$  values averaged across listeners. These values are close to zero until  $\rho=0.65$ , and did not rise significantly above zero until  $\rho=0.95$  in one-sample  $t$ -tests.<sup>2</sup>

### C. Discussion

#### 1. The extent of trading

As anticipated, the results show that spectrum level and interaural correlation trade against each other. This result indicates that changes in the central spectrum are sufficiently similar in perceptual effect to changes in the monaural power spectrum that features of each can become confused. For  $\rho = 0.5$ , it is clear from the data that trading is almost complete,

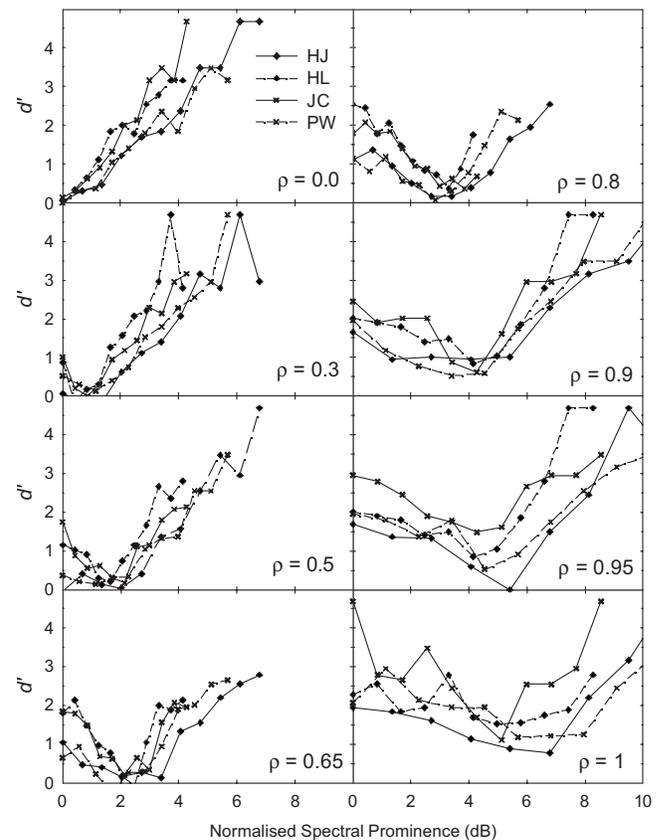


FIG. 4. Each panel shows psychometric functions for each of the four listeners at a specified value of  $\rho$ . There are 100 trials for each data point. The spectral prominence axis is normalized according to the regression slopes from Fig. 3 in order to align the  $d'$  minima.

because the stimuli are indiscriminable; performance close to the PSE never rises significantly above chance. Even at  $\rho = 0.65$  and  $0.8$  the minimum recorded value of  $d'$  is often close to zero. Although discrimination at the PSE improves at higher levels of  $\rho$ , this could be attributed to detection of differences in the perceived spatial extent or location of the target band and does not necessarily imply that loudness has not been equalized. Most remarkable is the fact that a V-shaped function is still observed at  $\rho=1$ , where the differences in spatial extent are quite prominent. Again, it seems likely that loudness was equalized in this case, but that listeners were able to differentiate the stimuli based on laterality in a proportion of trials; the limiting factor on their performance may have been their ability to ignore the loudness cue when it had been so reliable throughout the rest of the experiment.

The term “trading” is most commonly associated with conflicting interaural level differences (ILDs) and interaural time differences (ITDs). However, [Haftner and Carrier \(1972\)](#) showed that trading is never complete in that case. They used a same-different task rather than the odd-one-out technique used here, but they also transformed their results into  $d'$ , producing V-shaped psychometric functions for  $d'$  as a function of ILD for a series of ITD values. Their V-shaped functions never reached a minimum of zero. Indeed, even when it was assumed that listeners might perform the task using monaural level cues, the minimum observed  $d'$  was consis-

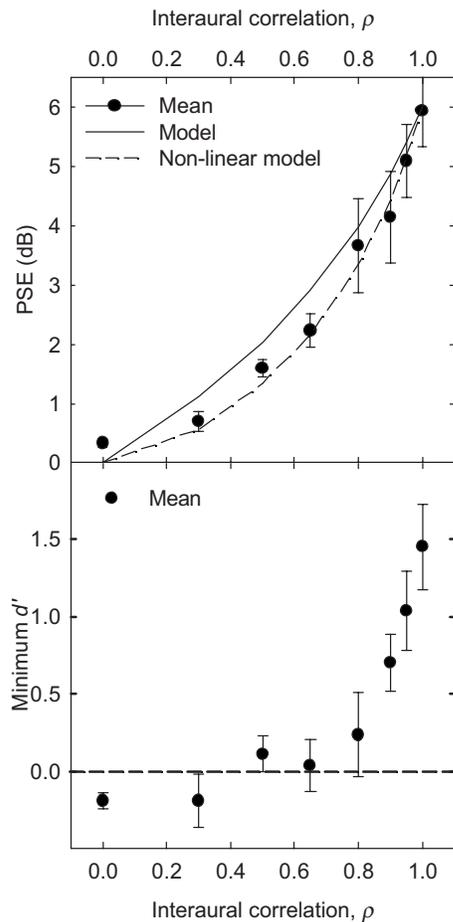


FIG. 5. The top panel shows the growth in the PSE (the spectral prominence in dB at the PSE) at each interaural correlation. The solid and dashed lines show predictions based on Eqs. (5) and (6) respectively (see Sec. II C). The bottom panel shows mean minima in  $d'$  (circles) taken from the bilinear fitted functions (see Fig. 2) for each listener as a function of interaural correlation,  $\rho$ .

tently higher than would be expected based on that cue alone. That was not the case in the current data. Listeners could have performed the task at the PSE using the available monaural cue of spectral prominence, which should have been quite adequate (e.g. 2–3 dB for  $\rho=0.5$ –0.8). They were, however, unable to ignore the contradictory information from the binaural cue. Despite this, the complete trading observed in the present data is not evidence that intensity and interaural correlation always trade perfectly. Nor does it show that they trade more effectively than ILD and ITD, because Hafter and Carrier trained their listeners very extensively, each completing at least 32,000 trials before commencing data collection. In comparison, listeners in the present experiment completed a total of 8,800 trials in the entire experiment. JC had extensive experience with similar experiments and HL had conducted pilot tests, but PW and HJ had no training. Indeed, it seems likely that some improvement would occur with more extensive practice. All listeners showed some improvements in score between the first and the second sets of 8 sessions (on average, performance increased from 72.1% to 76.7% correct).

The fact that listeners displayed a dip in  $d'$  even at  $\rho=1$  was surprising. For most of the conditions, the target

band is accessible via two cues. It appears louder than the flanking noise, and it is heard in a different location at one side or other of the listener's head. The location cue is largely nullified, however, by the fact that each non-signal interval also contains an interaurally uncorrelated target band. These bands are already strongly lateralized in the head, so the listener can only use this cue to perhaps detect less lateralization in the signal interval. Introspectively, the difference only became noticeable for  $\rho=1$ , since even  $\rho=0.95$  seemed to produce a strong sense of laterality. At  $\rho=1$  the relative absence of laterality was very obvious introspectively, yet all listeners continued to make the occasional mistake, presumably because they were perseverating to some extent in their use of the loudness cue. The apparent failure to exploit the laterality to the full in this case is quite fortunate, because the dip in performance allowed us to apply bilinear fits and derive PSEs from the  $\rho=1$  data and so to estimate the spectral change required to match the partial loudness caused by  $\rho=0$  in the non-signal intervals.

The PSEs at  $\rho=1$  averaged 6.1 dB, so an interaurally coherent target band needs to be 6.1 dB more intense than the interaurally incoherent target bands in the non-signal intervals for the stimuli to be maximally confusable. This suggests that the incoherence of the non-signal-interval target bands produces an increase in partial loudness equivalent to 6.1 dB. This figure is in reasonable correspondence with Gockel *et al.*'s (2009) value of 6.6 dB, and can be thought of as the equivalent dynamic range of the central spectrum at 500 Hz.

## 2. Relation to E-C theory

It would seem natural to interpret the effect of 6 dB in the central spectrum generated by the target band as equivalent to an increase in intensity of 6 dB over the flanking bands. This corresponds to a factor of 4, implying that uncorrelated noise has four times the potency of correlated noise in inducing loudness. The idea that interaurally uncorrelated sound is more potent than interaurally correlated sound is consistent with the observation that diotic spectral increments are perceptually insignificant near, and even substantially above, binaural detection threshold, whereas the addition of uncorrelated noise is much more easily detected (Culling, 2007). Since  $\rho$  is the proportion of correlated power, the total excitation of the target band,  $E_\rho$ , would, under this assumption, be proportional to the weighted sum of  $\rho$  and  $1-\rho$ ,

$$E_\rho \propto \rho + 4(1 - \rho) = 4 - 3\rho. \quad (3)$$

In equalization-cancellation (E-C) theory (Durlach, 1963, 1972), the internal representations of the stimulus at each ear,  $L(t)$  and  $R(t)$ , are first equalized in interaural time delay and then subtracted one from the other to produce a cancellation residue,  $Y(t)$ , which is used for detection in binaural unmasking, and also for detection of dichotic pitches (Durlach, 1962; Culling *et al.*, 1998a, 1998b). In terms of E-C theory, binaural loudness may thus be composed of three components, based on  $L(t)$ ,  $R(t)$  and  $Y(t)$ . Given that  $Y(t) \propto (1-\rho)$ , whereas  $L(t)$  and  $R(t)$  relate to the total noise level,

one can rearrange Eq. (3) to make the mathematically equivalent assumption that  $Y(t)$  is perceptually weighted three times more than the combination of the two monaural representations,  $L(t)$  and  $R(t)$ , and that the overall loudness is proportional to the sum of their weighted effects,<sup>3</sup>

$$E_\rho \propto 1 + 3(1 - \rho) = 4 - 3\rho. \quad (4)$$

The PSEs in the experiment can then be predicted from the ratio of the two excitation values for the signal and non-signal intervals, for which  $\rho=1$  and 0, respectively,

$$PSE_\rho = 10 \log_{10}\left(\frac{E_0}{E_\rho}\right) = 10 \log_{10}\left(\frac{4}{4 - 3\rho}\right). \quad (5)$$

This idea seems compatible with the overall pattern of data. Figure 5 (upper panel, solid lines) shows the PSEs predicted from Eq. (5). The model predictions correlate strongly with the data ( $r=0.985$ ). However, for the following reasons, this model is certainly too simplistic.

E-C theory is more usually applied to the binaural unmasking of signals. The strongest unmasking effect occurs when an out-of-phase signal is added to an in-phase noise (NoS $\pi$ ). If the signal were an out-of-phase noise band, it would generate a residue with twice the power of an uncorrelated noise band, so our simple model predicts that there should be a 9-dB binaural masking level difference for NoS $\pi$ . This prediction is somewhat less than observed difference, which is consistently in the range of 10–15 dB at 500 Hz (Durlach and Colburn, 1978). Above threshold, the model predicts that signals in NoSo and NoS $\pi$  should maintain equal loudness when they differ in level by 9 dB, but several studies have reported that tones in NoS $\pi$  grow more slowly in loudness than tones in NoSo (Hirsh and Pollack, 1948; Townsend and Goldstein, 1972; Soderquist and Shilling, 1990; Zwicker and Henning, 1991).

Part of this slow growth in loudness can be attributed to a saturating limit in binaural unmasking effects. Once the out-of-phase noise within a given frequency channel exceeds the in-phase noise (resulting in negative interaural correlation), E-C theory would predict that the equalization process will no longer be attracted to the correlated portion of the noise and so the out-of-phase portion (i.e., the signal) will begin to be cancelled instead. This happens at surprisingly low sensation levels; from Townsend and Goldstein's data, it occurs at about 8 dB SL.<sup>4</sup> Thus, signal loudness in NoS $\pi$  is predicted to grow in parallel with NoSo at low sensation levels, but sharply flatten off once the within-channel interaural correlation is reduced below zero. Such a plateau in loudness was observed by Culling *et al.* (2003) (expt 1., zero notch width), who measured cumulative  $d'$  for loudness discrimination across the full range of noise-band interaural correlations from  $-1$  to  $1$ . Nonetheless, Townsend and Goldstein observed a substantial decline in the difference in signal level that yielded equal signal loudness at only 5 dB SL, indicating a genuine decline in sensitivity to  $Y(t)$  as it increased in magnitude. Declining sensitivity to  $Y(t)$  as it increases implies greater sensitivity to  $\rho$  as it approaches 1, so we can apply such a model by adding an expansive non-

linearity to  $\rho$  in Eq. (5). A satisfactory fit was found when an exponent of 1.5 [Eq. (6)] was applied, producing a high correlation ( $r=0.997$ ) with the observed PSEs (Fig. 5, dashed line),

$$PSE_\rho = 10 \log_{10}\left(\frac{4}{4 - 3\rho^{1.5}}\right). \quad (6)$$

### 3. Across-frequency effects

The salience of uncorrelated or out-of-phase signals also seems to depend critically on the context of presentation. Edmonds and Culling (2009) recently reported an effect of interaural correlation on loudness in which the interaural correlation of the entire stimulus spectrum was manipulated. Edmonds and Culling found very little difference in loudness between narrow bands of correlated and “anticorrelated” (out-of-phase) noise. This makes sense in terms of E-C theory, because the phase difference in the anticorrelated noise would be largely compensated in the equalization process, enabling the subsequent cancellation to largely eliminate this noise from  $Y(t)$ . For anticorrelated noise to be salient, it needs to be mixed with at least the same level of correlated noise to create a correlation closer to zero. Correlated and anticorrelated noise might be added at the same frequencies in the stimulus, or may mix as a result of entering the same frequency channel. If the majority of the noise in a frequency channel is correlated, the E-C process will remove that portion of the noise and all of the anticorrelated noise will contribute to  $Y(t)$ . Consistent with these ideas, Edmonds and Culling found that uncorrelated noise (equivalent to a 50/50 mix of correlated and anticorrelated noise) was louder than correlated noise. However, uncorrelated noise was matched in loudness to correlated noise that was only about 2 dB more intense. Edmonds and Culling remarked that this effect seemed small compared to the cumulative  $d'$  values from Culling *et al.* (2001), which, as noted above, could be as large as 6. The present data use a more directly comparable measurement technique, and confirm that listeners are more sensitive to  $\rho$  than observed by Edmonds and Culling when variations in  $\rho$  are presented in a small part of the spectrum that is otherwise filled with diotic noise. Across-frequency variation in  $\rho$  would thus appear responsible for the larger effect found here.

It is difficult to see why such across frequency variations should have such a strong effect on the loudness. One possibility is that the lateral inhibition mechanism suggested by Klein and Hartmann (1981) to explain the binaural edge pitch increases the perceptual salience of narrow bands of uncorrelated noise. Although the binaural edge pitch can be explained without recourse to lateral inhibition (Culling *et al.*, 1998a), it appears to be essential to explain binaural coherence edge pitch (Hartmann and McMillon, 2001). Further experiments will be required to ascertain whether this mechanism is sufficient to explain our contrasting findings.

### III. CONCLUSIONS

The profile of interaural coherence as a function of frequency appears to evoke the perception of a central spectrum

analogous to the monaural perception of the frequency power spectrum. This central spectrum has a dynamic range that is perceptually equivalent to about 6 dB of the monaural range. Spectral prominences in the central spectrum are perceived sufficiently similarly to those in the power spectrum that these monaural and binaural features can be conflated in the overall perception of the stimulus.

## ACKNOWLEDGMENTS

We would like to thank Katrin Krumbholz and two anonymous reviewers for valuable comments on earlier versions of the manuscript.

<sup>1</sup>Where that maximum occurs at zero delay, it has the same value as the interaural correlation. As a result, these terms can be used interchangeably for many stimulus designs. Incoherence refers to reductions in these values.

<sup>2</sup>Although the error bars do not overlap with zero at  $\rho=0.9$ , the result was non-significant, because, with  $df=3$ , the critical value of  $t$  was quite large. No Bonferroni correction was employed; this would have reduced sensitivity further.

<sup>3</sup>The difference between Eqs. (3) and (4) is between separating portions of the stimulus (the correlated and uncorrelated noise) in Eq. (3) and separating the two processing pathways (monaural and binaural) in Eq. (4). The uncorrelated noise contributes to both monaural and binaural excitation in Eq. (4), but received less weight for its binaural contribution than in Eq. (3).

<sup>4</sup>This value was obtained by measuring the signal level in NoS $\pi$  at which zero interaural correlation was observed at the output of a matched pair of gammatone filters centered on the signal frequency of 500 Hz. The threshold signal level in NoSo was taken from Townsend and Goldstein's data at 500 Hz in 40 dB/Hz noise.

Akeroyd, M. A., Moore, B. C. J., and Moore, G. A. (2001). "Melody recognition using three types of dichotic-pitch stimulus," *J. Acoust. Soc. Am.* **110**, 1498–1504.

Akeroyd, M. A., and Summerfield, A. Q. (2000). "The lateralization of simple dichotic pitches," *J. Acoust. Soc. Am.* **108**, 316–334.

Bilsen, F. A. (1977). "Pitch of noise signals: Evidence for a 'central spectrum'," *J. Acoust. Soc. Am.* **61**, 150–161.

Cramer, E. M., and Huggins, W. H. (1958). "Creation of pitch through binaural interaction," *J. Acoust. Soc. Am.* **30**, 413–417.

Culling, J. F. (1999). "The existence region of Huggins' pitch," *Hear. Res.* **127**, 143–148.

Culling, J. F. (2007). "Evidence specifically favoring the equalization-cancellation theory of binaural unmasking," *J. Acoust. Soc. Am.* **122**, 2803–2813.

Culling, J. F., Colburn, H. S., and Spurchise M. (2001). "Interaural correlation sensitivity," *J. Acoust. Soc. Am.* **110**, 1020–1029.

Culling, J. F., Hodder, K. I., and Colburn, H. S. (2003). "Interaural correlation discrimination with spectrally-remote flanking noise: Constraints for models of binaural unmasking," *Acta. Acust. Acust.* **89**, 1049–1058.

Culling, J. F., Marshall, D. H., and Summerfield, Q. (1998a). "Dichotic pitches as illusions of binaural unmasking. II. The Fourcin pitch and the dichotic repetition pitch," *J. Acoust. Soc. Am.* **103**, 3527–3539.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (1998b). "Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the 'binaural edge pitch'," *J. Acoust. Soc. Am.* **103**, 3509–3526.

Durlach, N. (1972). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 369–462.

Durlach, N. I. (1962). "Note on the creation of pitch through binaural interaction," *J. Acoust. Soc. Am.* **34**, 1096–1099.

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

Durlach, N. I., and Colburn, H. S. (1978). "Binaural phenomena," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. **IV**, pp. 365–466.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (1986). "Interaural correlation discrimination: II. Relation to binaural unmasking," *J. Acoust. Soc. Am.* **79**, 1548–1557.

Edmonds, B. A., and Culling, J. F. (2009). "Interaural correlation and the binaural summation of loudness," *J. Acoust. Soc. Am.* **125**, 3865–3870.

Gockel, H. E., Carlyon, R. P., and Plack, C. J. (2009). "Pitch discrimination interference between binaural and monaural or diotic pitches," *J. Acoust. Soc. Am.* **126**, 281–290.

Haftner, E. R., and Carrier, S. C. (1972). "Binaural interaction in low-frequency stimuli: The inability to trade time and intensity completely," *J. Acoust. Soc. Am.* **51**, 1852–1862.

Hartmann, W. M., and McMillon, C. D. (2001). "Binaural coherence edge pitch," *J. Acoust. Soc. Am.* **109**, 294–305.

Hirsh, I. J. (1948). "The influence of interaural phase on interaural summation and inhibition," *J. Acoust. Soc. Am.* **20**, 536–544.

Hirsh, I. J., and Burgeat, M. (1958). "Binaural effects in remote masking," *J. Acoust. Soc. Am.* **30**, 827–832.

Hirsh, I. J., and Pollack, I. (1948). "The role of interaural phase in loudness," *J. Acoust. Soc. Am.* **20**, 761–766.

Klein, M. A., and Hartmann, W. M. (1981). "Binaural edge pitch," *J. Acoust. Soc. Am.* **70**, 51–61.

Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.

Moore, B. C. J., and Glasberg, B. R. (2007). "Modelling binaural loudness," *J. Acoust. Soc. Am.* **121**, 1604–1612.

Raatgever, J., and Bilsen, F. A. (1977). "Lateralization and dichotic pitch as a result of spectral pattern recognition," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 443–453.

Smith, J. E. K. (1982). "Simple algorithms for M-alternative forced-choice calculations," *Percept. Psychophys.* **31**, 95–96.

Soderquist, D. R., and Shilling, R. D. (1990). "Loudness and the binaural masking level difference," *Bull. Psychon. Soc.* **28**, 553–555.

Townsend, T. H., and Goldstein, D. P. (1972). "Suprathreshold binaural unmasking," *J. Acoust. Soc. Am.* **51**, 621–624.

Zhang, P. X., and Hartmann, W. M. (2008). "Lateralization of Huggins pitch," *J. Acoust. Soc. Am.* **124**, 3873–3887.

Zwicker, E., and Henning, G. B. (1991). "On the effect of interaural phase differences on loudness," *Hear. Res.* **53**, 141–152.