Interaural Correlation Discrimination

Culling, John F[1]; Colburn, H. Steven[2]; Hodder, Kathryn. I[1]

[1]Sch. of Psychology, Cardiff University; [2]Dept. Biomedical Engineering, Boston University.

[1]P.O. Box 901 Cardiff CF10 3YG. UK. (Tel. +44 29 2087 4523; FAX +44 29 2087 4858; email CullingJ@cf.ac.uk)

[2]44, Cummington St. Boston. MA 02215. USA. (Tel +1 617 353 4342; FAX +1 617 353 6766; email: colburn@bu.edu)

ABSTRACT

It is difficult to discriminate experimentally between different models of binaural unmasking, because the internal representations used in these models always display changes when the stimulus parameters are changed; if listeners are sensitive to these parameters, each model can account for that sensitivity. Reversing this logic, our experiment shows that listeners can lack sensitivity to changes in interaural correlation in circumstances where most models would predict them to be sensitive. Listeners show maximum sensitivity to the interaural correlation of a band of noise at zero correlation when remote flanking bands of noise are present.

INTRODUCTION

When sounds come from different directions, a listener with normal hearing can detect or identify the sounds more easily than if they come from the same direction. A component of this ability is dependent upon the differences in arrival time of the two signals at each ear. This component is illustrated by the phenomenon of binaural unmasking (Hirsh, 1948); if the interaural phase of a tonal signal differs from that of a broadband noise masker, a normally-hearing listener's masked detection threshold is reduced by up to 15 dB.

Most models of binaural unmasking are at least loosely based on the architecture proposed by Jeffress (1948) for sound localisation. In Jeffress' model, axons arising from corresponding places in each cochlea converge through a ladder-like network of neural delay lines onto coincidence detecting neurons. If the internal delay generated by the difference in transmission time from each ear is equal and opposite to the difference in the signal's arrival time at each ear, the two delays will cancel, bringing the series of action potentials from each ear into register. Only the coincidence-detecting neurons with this internal delay will therefore fire, indicating the direction of sound in the left/right dimension. Jeffress envisaged a two-dimensional matrix of coincidence detectors with different characteristic frequencies (by virtue of connection to different places in the two cochleae) and different characteristic delays (by virtue of varying lengths of axon). Support for Jeffress' model comes from the discovery of "EE" units in the auditory brainstem (Goldberg and Brown, 1969), which are selectively excited by sounds of the appropriate frequency and interaural delay. Goldberg and Brown also found "EI" units, which are selectively inhibited by signals with a particular interaural delay.

Jeffress' model can be implemented computationally by cross-correlating the corresponding frequency channels of a pair of filterbanks designed to simulate cochlear filtering at each ear. Each cross-correlation function represents the activity pattern across an array of EE units with different characteristic delays, but the same characteristic frequency. Culling and Summerfield (1995) created a crude simulation of the response of a similar array of EI units by replacing the multiplication operation in the cross-correlation with an absolute difference operation, a process that one might call "cross-cancellation." The pattern of response predicted by such models changes when the interaural phase of a signal differs from that of a masking noise. In particular, the cross-correlation function for the combined stimulus will have a peak that is less than unity. This fact has lead many researchers to think about binaural unmasking in terms of detection of a reduction in interaural correlation at the signal frequency. However, the existence of *any* changes in pattern between detectable signals and undetectable signals is regarded as evidence in favour of the models. This approach is most explicit in the models of Colburn (1973, 1977), based on an EE-type array and Breebaart *et al.* (2001a,b,c), based on an EI-type array. These models base detection on any differences in response pattern for masker alone and masker+signal.

Since different models are structurally similar and all produce changes in activity pattern when any change to the interaural parameters of the stimulus is introduced, it is difficult to discriminate between competing models experimentally. However, the experiment presented in this paper is an attempt to produce such discriminating data. The problem of all models producing changes in their pattern and therefore all predicting detection of any change in the binaural configuration of the stimulus is avoided by looking for situations in which the pattern changes, but the listeners are *insensitive* to the change.

The approach is based upon the mE-C model of Culling and Summerfield (1995) and the experimental methodology of Culling, Colburn and Spurchise (2001). A distinctive feature of Culling and Summerfield's model is that rather than incorporating all information from an array of EI units, it pre-selects a subset of that information, and assumes that only that subset is accessible to the listener. Briefly, the model scans across delay at each frequency, for the minimum in the cross-cancellation function. The value of the minimum at each frequency forms one point in a "recovered spectrum." The listener has access only to the recovered spectrum. An attractive feature of the model is that tones whose interaural phase differs from that of the noise, show up as peaks in the recovered spectrum. Effectively, the model detects frequency channels in which the stimulus is interaurally uncorrelated. However, the model is blind to the delay at which information relating to the signal was detected. A disadvantage of this scheme is that the model cannot, therefore, predict the perceived locations of tones in noise or of a group of binaural phenomena known as dichotic pitches (Bilsen, 1977; Raatgever and Bilsen 1986). Listeners do perceive distinct locations for these sounds, which are left unexplained. From the present standpoint, this feature is useful, because it predicts that a well-modulated cross-cancellation function will never be perceptually salient, because it will have low minima, regardless of the delay at which those minima occur or of the delays at which minima occur at other frequencies. In other words, according to this model (and in contrast to other models) listeners are predicted to be unable to discriminate stimuli in which the delay of the minimum is changed, even quite dramatically.

The methodology used here is an extension of that used by Culling, Colburn and Spurchise (2001). They were interested in the ability of the binaural system to encode the intensity of a signal. They reasoned that as the intensity of a signal increases, the correlation at the signal frequency progressively decreases. Therefore, such decorrelation might be used by the auditory system as a direct perceptual surrogate for signal intensity. In order to test this hypothesis, they controlled the interaural correlation of a sub-band within a broadband noise. All other frequencies had an interaural correlation of unity. As expected, listeners heard the sub-band standing out from the rest of the noise when its correlation was reduced. This percept is a form of dichotic pitch. As the correlation became lower the sub-band became more prominent; it was perceived as louder. Using a loudness discrimination paradigm, Culling *et al.* derived a slope of increasing loudness as a function of correlation between 1 and 0 correlation.

Culling *et al.* only used target correlations (ρ_T) between 0 and 1, because the predictions become problematic for negative correlation. Two effects are at work. First, since the cross-

correlation function of a narrow band of noise is approximately sinusoidal, if the correlation in the target sub-band is negative at zero delay, then it is positive at some other delay. If the binaural system is sensitive to correlation *per se* (as predicted by the mE-C model) then one might expect the negative correlation to be *less* detectable. Second, frustrating this prediction, if the correlation in the target band is negative, it will mix, at the edges of this band, with the correlated noise. This mixture will tend to be uncorrelated and thus highly detectable if the binaural system is sensitive to decorrelation. The aim of the present experiment is to control the second effect in order to reveal the first. The second effect was attenuated by introducing spectral notches between the target and spectrally-flanking noise (see figure 1).
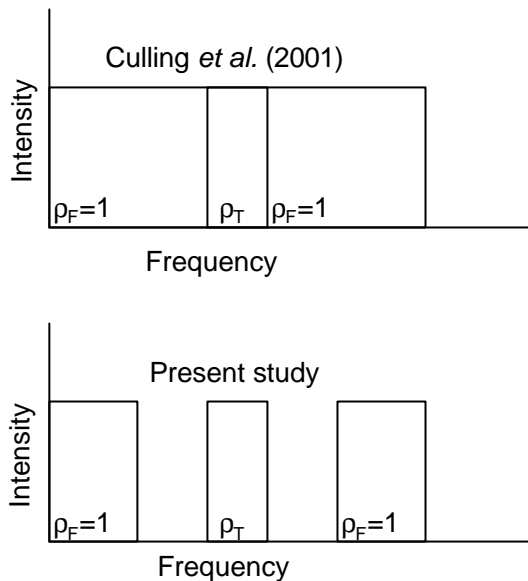
The flanking bands were not eliminated entirely, because if the target band were presented alone, its perceived width and location would change with its interaural

Figure 1. Schematic illustrations of the stimuli used by Culling *et al*. (2001) and in the present study.

correlation. Since the study is motivated by questions about binaural unmasking, it is undesirable to allow any potentially confounding cue to be available to the listeners. The flanking bands reduce changes in localisation, because sound localisation is an integrative process that includes information from all frequencies. The flanking bands contain more energy, so they dominate the perceived location of the stimulus as a whole and minimise any changes due to the changing correlation of the target band.

METHOD

The discriminability index, *d'*, was measured for various pairs of ρ_T values. Individual *d'* values were measured by presenting a block of 50 two-interval trials. In each trial, the listener heard, in a random order, two stimuli, one with each of the two different values of ρ_T under test. Listeners identified the stimulus for which the target band sounded louder by typing '1' or '2' on a computer terminal. No feedback was given. *d'* for each block was calculated, always counting identification of the lower correlation value as a 'hit.' If the higher correlation was heard as louder by the listener, a negative *d'* would therefore result.

Stimuli were generated online, using a TDT array processor, and presented to listeners through a TDT psychoacoustic rig (DD1; PA4; HB6) and Sennheisser HD590 headphones in a single-walled IAC sound-attenuating chamber, situated within a sound treated room. The flanking bands had a correlation, ρ_F, of 1 or of −1. The target bands had correlations from the set (1.0, 0.8, 0.5, 0, -0.5, -0.8 and −1.0). Each value was compared with the two subsequent values in this series, making 11 measurements in all. As in Culling *et al.*'s study, the correlation of the target band was set precisely to the desired correlation value over its 400-ms duration by using the Gramm-Schmidt orthogonalisation procedure.

Target bands were 1 ERB wide (≈80 Hz) and centred on an ERB scale (Moore and Glasberg, 1983) around 500 Hz. In a given experimental session, the spectral notches, in either side of the target bands, were 0, 1, 2, or 3 ERBs wide (around 0, 75, 155 and 235 Hz on average). The zero notch width condition was equivalent to the stimuli used by Culling *et al.* (2001). Each listener completed two sessions at each spectral notch width and for each flanking-band correlation. The order of the 11 measurements was counterbalanced across these two sessions. With two, flanking band correlations, and four notch widths, this required a total of 16

one-hour sessions. During each block, subjects were periodically reminded of the target band, by playing this band diotically and in isolation.

RESULTS

As in Culling *et al.* (2001), sensitivity functions to interaural decorrelation were derived using cumulative *d'* (Green and Swets, 1967). A fourth-order polynomial function was fitted to these individual measurements in order to produce a continuous sensitivity function. These functions were constrained to pass through the origin when the correlation of the target band matched that of the flanking bands, on the principle that when ρ_T and ρ_F are identical there should be no reason for the target band to stand out. The number of free parameters in the function was thereby reduced to three.

Figure 2 shows an example of how the raw *d'* values contributed to the derivation of a cumulative *d'* curve for listener DC using a notch width of 2 ERBs. The fitting algorithm adjusted the function parameters to minimise the difference between the measured *d'* values and the vertical change in the cumulative *d'* curve

Figure 2. Illustration of the process of fitting cumulative *d'* curves (thick lines) to raw *d'* measurements (thin lines).

over the same range of ρ_T. Thin lines in the figure show the raw *d'* measurements. They are adjusted in vertical position so as to fall on the fitted curve. The dotted line at $\rho_T=0$ emphasises the fact that the fitted curve reaches a peak at or around $\rho_T=0$. In terms of the raw data, it can be seen that this effect is the result of *d'* values on either side of $\rho_T=0$ having opposite sign (i.e. the lines representing the raw data slope in opposite directions). In terms of listeners' responses, this means, for instance, that when comparing $\rho_T=0.5$ and $\rho_T=0.8$, the listener reported that 0.5 sounded louder, but when comparing $\rho_T=-0.5$ and $\rho_T=-0.8$, -0.5 was louder.

Figure 3 shows only the fitted functions for three subjects individually for both $\rho_F=-1$ (left panel) and $\rho_F=1$ (right panel). Each panel shows curves for different widths of spectral notch about the target band. When the spectral notch is wide there is a clear peak in each function, centred at or around $\rho_T = 0$. However, when the spectral notch is narrow, or absent (solid lines), the function tends, instead, to plateau or to continue rising beyond $\rho_T=0$.

DISCUSSION

The results indicate that zero correlation is the most perceptually salient value in terms of the perceived loudness of the target band. Where the target and flanking bands were separated by a spectral notch the cumulative *d'* function peaks at or around this value of correlation. It is important to note that in many cases the target band's correlation contrasted with that of the flanking band, yet it was not heard as particularly prominent as a result of this contrast.

The results of this experiment favour the hypothesis that the binaural system detects frequency channels that are uncorrelated and that this detection of interaural decorrelation operates independently in different frequency channels. The maximally loud target bands are thus heard at $\rho_T=0$, regardless of the correlation elsewhere in the stimulus. However, this interpretation should be qualified by the observation that a simple application of this hypothesis would predict that the cumulative *d'* curves should be symmetrical about $\rho_T=0$. Even when the spectral notches are wide, this is seldom the case. One can account for the deviation from perfect symmetry by considering two additional factors.
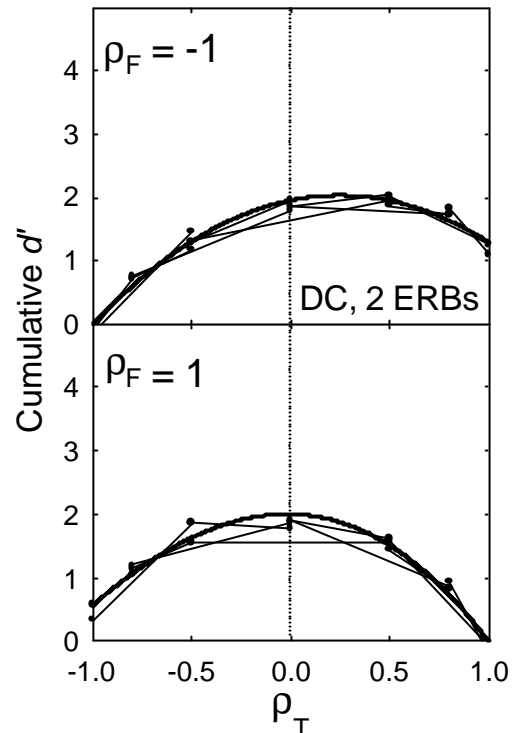
First, the limited spectral resolution in the auditory system means that, even with a spectral notch 3 ERBs wide, one cannot eliminate entirely the mixing of noise from the target and flanking bands. Where such mixing occurs, the frequency channels centred on the notch may receive a mixture of correlated and anti-correlated noise. This noise will, when mixed equally, be uncorrelated, and therefore salient to the binaural system. Such an effect can explain the fact that few of the curves in Figure 3 return to zero cumulative d' at the central axis of the figure, where such contrasts between target and flanking noise occur.

Second, while the interaural correlation of a band of anti-correlated noise is positive at delays well away from zero, it is never unity. Anti-correlated noise should, therefore, be predicted only to be less salient than uncorrelated noise, and not to be undetectable. This effect is relatively subtle, but is obvious in the following manifestation. Where the flanking bands are anti-correlated, and therefore somewhat salient, detection of decorrelation in the target band is in competition with the flanking bands. This effect should produce a skew in Figure 3, such that higher d' values are observed in the right hand panels. This was quite strongly the case for listener IS, but less so for other listeners.

In terms of the competition between different models, the present results are most consistent with the mE-C model of Culling and Summerfield (1995). This model detects decorrelation within frequency channels and is blind to the delays at which the decorrelation occurs. Except where the spectral notches are absent, Figure 3 indicates that uncorrelated target bands are always more salient.
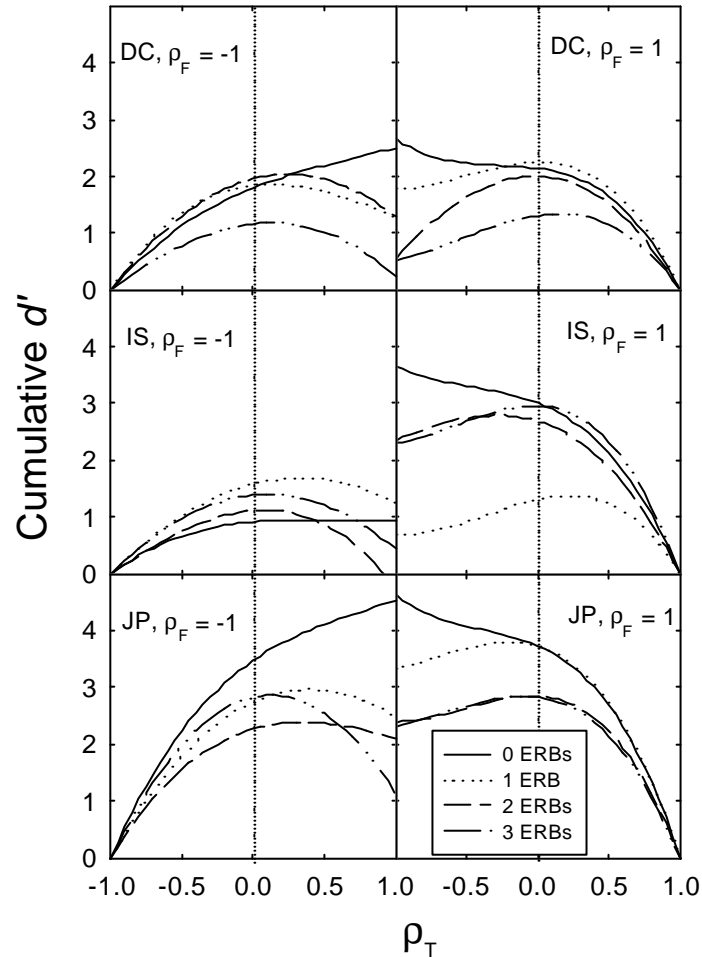

Figure 3. Fitted cumulative d' curves for three participants using correlated and anticorrelated flanking bands (ρ_F=1,-1) and three widths of spectral notch (0,1,2,3 ERBs).

The models of Colburn (1973, 1977) and of Breebaart *et al.* (2001a,b,c) predict that listeners can detect any change in the activity of an EE or EI matrix that exceeds some threshold. It is not clear how such detection should be interpreted strictly in term of the perceived loudness that was measured, here. However, in terms of the loudness, the present data suggest that some quite large changes are undetectable. Where the cumulative d' curve has a peak around ρ_T=0, values of ρ_T on either side of zero will have equal cumulative d'. Some such comparisons were made directly; looking at the bottom panel of Figure 2, among the line segments that represent the raw data is a line between ρ_T=0.5 and ρ_T=-0.5. The response of an array of EE or EI neurons should be quite different for these two stimuli, yet this listener could not discriminate their loudness. The cumulative d' curves predict that one should be able to observe such insensitivity with more extreme comparisons. For instance, the same panel of Figure 2, indicates that ρ_T=0.9 should sound the same as ρ_T=-1. To test the Colburn and Breebaart

models more directly, we plan to demonstrate similar insensitivity using a 3-interval, odd-one-out task, which would allow listeners to employ any cue to which they have access.

The present results also seem inconsistent with Raatgever and Bilsen's (1986) CAP model. Their model assumes that when listeners attend to a particular direction, they pick out one internal delay in an EE-based matrix. Scanning across frequency at that delay, the model recovers a spectrum that is thought to reflect sounds coming from the corresponding direction. Further, recovered spectra featuring prominent peaks are thought to draw the listener's attention. The use of across-frequency scans makes this model sensitive to across-frequency contrasts in correlation. A prediction of the CAP model for the present stimuli is that correlated bands with anti-correlated flanking bands should be maximally prominent, since the correlation will produce a peak at zero delay, whereas the flanking bands will have troughs at the same delay. Assuming that such prominence would be reflected in the perception of a loud target band, one would expect the CAP model to predict the maximum cumulative *d'* for this stimulus. However, except where the spectral notches are absent from the stimuli, the data show that such stimuli are consistently less salient than those in which the target band is uncorrelated.

CONCLUSIONS

Listeners perceive noise with zero interaural correlation as louder than noise of equal intensity with any other correlation. The perception of loudness at each frequency is largely independent of the correlation at other frequencies. This pattern of sensitivity is consistent with the mE-C model of Culling and Summerfield (1995), and inconsistent with most other models.

REFERENCES

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

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001a). 'Binaural processing model based on contralateral inhibition. I. Model structure" *J. Acoust. Soc. Am.* **110**, 1074-1089.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001b). 'Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters" *J. Acoust. Soc. Am.* **110**, 1089-1105.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001c). 'Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters" *J. Acoust. Soc. Am.* **110**, 1105-1118.

Colburn, H. S. (1973). "Theory of binaural interaction I: General strategy and preliminary results on interaural discrimination" J. Acoust. Soc. Am. **54**, 1458-1470.

Colburn, H. S. (1973). "Theory of binaural interaction II: Detection of tones in noise" *J. Acoust. Soc. Am.* **61**, 525-533.

Culling, J. F. and Summerfield, Q. (1995). "Perceptual segregation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay," *J. Acoust. Soc. Am.* **98**, 785-797.

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

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

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* (Wiley, New York)

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

Jeffress, L.. A. (1948). "A place theory of sound localization" *J. Comp. Physiol. Psych.* **61**, 468-486.

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.

Raatgever, J. and Bilsen, F. A. (1986). "A central theory of binaural processing: Evidence from dichotic pitch" *J. Acoust. Soc. Am.* **80**, 429-441.