Theme 3: Binaural Interactions

Binaural interactions of sound stimuli, *made possible* by the two inputs to the auditory system physically displaced from each other, enable many species to localize sounds in space. For human, binaural hearing also provides a mechanism to improve selective attention to speech in a noisy environment (Freyman *et al*, 1999; Ebata, 2003). *[not only speech and not only humans]* However, the locations of sound sources are not mapped directly at the sensory periphery (Macpherson *et al*, 2002). Instead, locations must be derived by acoustical cues such as interaural time difference (ITD), interaural level difference (ILD) and spectral cues. These cues are the result from incident sound waves interacting with the external ears, head and upper torso. Human and barn owls *[and cats]* can resolve ITD near or less than ten microseconds (Grothe, 2003). Considering that the duration of an action potential is *at least one* order of magnitude greater, the exact mechanism in neural coding for such fine timing information is of great research interest and in this theme paper, some of the latest findings are reported and the implication to the validity of the long-standing Jeffress coincidence model is discussed. Localization of sound can also be influenced in an echoic environment. Physiological recordings of different neural units at different level of the ascending auditory pathway as well as psychophysical studies performed on animals are also documented in this report, addressing some important behavioral phenomena, such as the "precedence effect", relevant to such echoic surrounds. Literature on the duplex theory on sound localization, proposed by Lord Rayleigh, are also reviewed to determine the relevance of this century-old binaural theory in light of data obtained from modern psychophysical experiments.

British physicist Lord Rayleigh postulated a century ago that the primary cue to localize lateral positions of low-frequency tonal stimulus (< 500 Hz) is ITD, while ILD is used for high-frequency tonal sources, based on the observations of acoustic shadowing effects by the head. This dichotomy for localization of tones in the azimuth has come to be known as the "Duplex Theory". Macpherson *et al* (2002) reexamined this theory to ascertain whether the theory is applicable to localization of naturally occurring broadband sounds, simulated by using virtual auditory space (VAS) techniques. As part of the experimental design, the investigators manipulated the stimuli by imposing a whole-waveform delay or attenuation to the signal at one of the ears. The imposed bias and the observed bias in terms of ITD and ILD are then expressed as a ratio to represent the salience, or the weight, of that cue[1] (see Figure 1). One of the ambiguities in interpreting the weights of these cues using the methods described above is that the disparity between the observed and imposed bias may be due to the influence of other cues, or it can simply be attributed to undershooting of the subjective response. Such phenomenon observed in psychophysical experiments is frequently documented, and it will be further discussed in the later part of this report (c.f., Tollin *et al*, 2003).

[1] Therefore, a cue weight close to zero signifies a very weak cue, while if the observed bias equals the imposed bias, the cue weight will be unity.

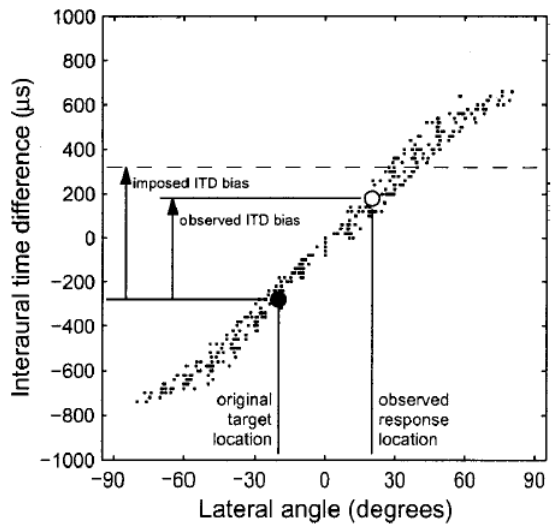

Figure 1 Computation of observed cue bias. (Macpherson *et al*, 2002, Figure 3)

Two experiments were carried out to measure the ITD- and ILD-bias weights. Figure 2 (Left) summarizes the findings of Macpherson and his co-author. From the figure, it can be seen that the ITD weights are larger than the ILD weights for low-pass targets and the reverse for high-pass targets. There are moderate to large weights on both ITD and ILD for wideband targets, although ITD was usually weighted more strongly. The authors hypothesized that 'low frequency ILDs have little effect on apparent source direction because they are reserved as cues for near-field distance perception'. In another experiment, based on the results of other lateralization studies showing that listeners are sensitive to ITDs in high-frequency complex sounds, the investigators manipulated the onset and ongoing envelope-based ITD cues in high-pass noise to explore the relative influence of ITD cues of such stimulus. From Figure 2 (Right), it can be seen that both onset and ongoing envelope ITD cues contributed to the relatively minor role of high-frequency ITD information. The authors, however, noted that the greatest high-frequency ITD weights were observed for listeners who were able to make use of ongoing cues.

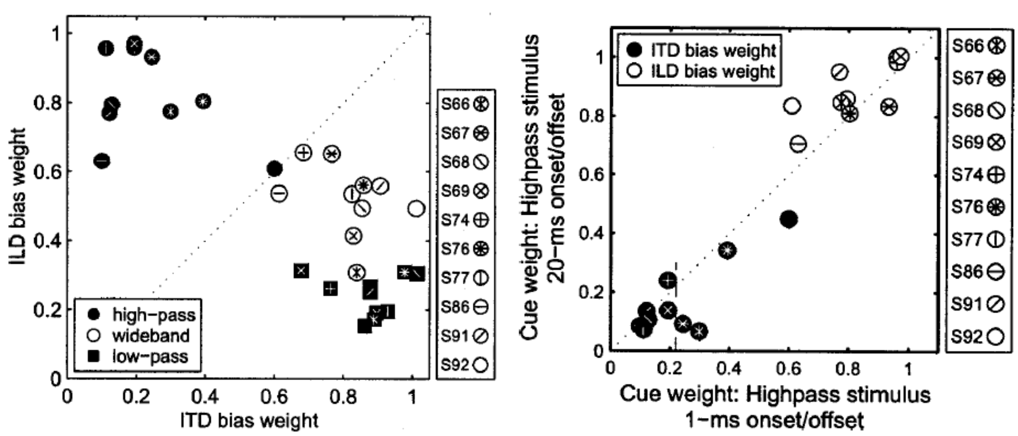

Figure 2 (Left) Measured ITD- and ILD-bias weights. (Right) Effect on onset ramp duration on high-pass ITD and ILD bias weights. The long onsets were intended to reduce the salience of the onset ITD cue, and the amplitude modulation was intended to enhance the salience of the ongoing envelope ITD cue. (Macpherson *et al*, 2002, Figure 5, 6)

Macpherson *et al* (2003) also manipulated the interaural level spectra (ILS, i.e., patterns of ILD across frequency) to explore its importance for sound localization. They attempted to bias the perceived location of virtual free-field targets without introducing unnatural patterns of ILD across the frequency spectrum. From the results, the authors suggested that the detailed shape of ILS was not as effective a cue as a net ILD bias. Furthermore, in another experiment, they investigated the importance of monaural spectral cues for lateral angle localization. The authors found that monaural spectral cues had little or no influence on perceived lateral angle.

The fine structure as well as the envelope of a waveform conveys ITD information. Bernstein (2001) used "transposed" stimuli to find out the relative ITD sensitivity compared to sinusoidally amplitude-modulated (SAM) stimuli with a high-frequency carrier, and a low-frequency tone, with its frequency being equivalent to the envelope *rate* of the aforementioned stimuli. The generation of the "transposed" stimuli entails multiplying a half-wave-rectified, low-frequency tone by a high-frequency sinusoidal carrier. From Figure 3, it can be seen that after the assumed peripheral processing[2], the output of the low-frequency tonal stimulus is identical to that of the "transposed" stimulus, while output of the same peripheral processing for the SAM stimulus has less distinct peaks. The author argues that it is reasonable to assume that the corresponding neural discharge patterns for the processed tone and the "transposed" stimuli would be less dispersed in time than the SAM tone. However, it is important to note that for the tonal stimulus, there will be a greater depth in the spike rate modulation compared to the "transposed" stimulus, and thus even though in Figure 3, the output characteristics look identical, their neural discharge patterns do differ.

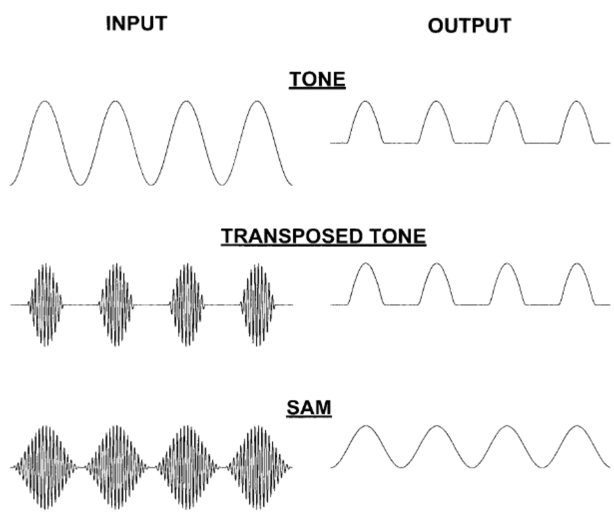

Figure 3 Different input stimuli processed by the same band-pass filtering, rectification and low-pass filtering. Note the output for tonal and "transposed tone" stimulus are equivalent. (Bernstein, 2001, Figure 8)

In Bernsteins' experiment, he showed that the threshold ITDs obtained with the transposed stimuli were uniformly smaller than those obtained with the SAM tones (Figure 4). The author also argues that since the thresholds for the high-frequency "transposed" stimuli are very similar to that of the low-frequency pure tones for frequencies of 128 Hz and 256 Hz, the ability to discriminate changes in ITD using the "transposed" envelope comes very close to that measured with low-frequency stimulation. Furthermore, in another experiment, Bernstein showed that low-frequency noise results in greater extents of laterality than the high-frequency SAM tone, and the high-frequency "transposed" tone even exceeds the low-frequency noise in terms of laterality potency. The authors view these results as strong evidence to support the hypothesis[3] that the difference in ITD potency

[2] I.e., band-pass filtered, half-wave rectified and then low-pass filtered.

[3] Hypothesis first postulated by Colburn and Esquissaud (1976).

typically observed for low and high frequencies results from differences in the peripheral coding processes rather than from more central binaural mechanisms.

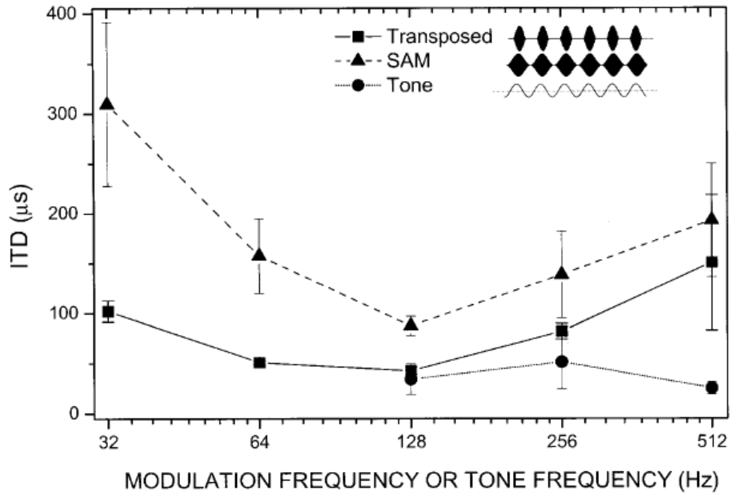

Figure 4 Threshold ITDs as a function of the modulation or pure-tone frequency. (Bernstein, 2001 Figure 9 from Bernstein and Trahiotis, 2001)

In reverberant environments, other spatial perceptual phenomena, collectively known as the precedence effect (PE), are required for sound location in the face of conflicting reflections acoustical cues. The PE has been characterized in humans previously, but given there was very little evidence that the animals used in the physiological studies experience the same auditory phenomenon, Tollin *et al* (2003) assessed the localization capabilities of trained cats, derived by their saccadic eye movements. When the subject was presented two sounds, a leading sound and a delayed sound simulating a single reflection temporally separated by the interstimulus delay (ISD) variable, the apparent location was recorded, and this experiment was carried out both on the horizontal and the vertical plane. Figure 5 shows the results of the azimuth response. When ISD is below 1 ms, there is evidence of summing localization[4]. Localization dominance[5] is observed for these cats, and similar to human data, there is also some undershoot in their responded location. The authors postulate that the statistically significant undershoots can be a real, but small, residual effect of the lagging stimulus “pulling” the apparent location. However, they also indicate that the degree of undershooting can be reduced by freeing the head of the cats. These cats, however, do not experience summing localization with these paired sources in the vertical plane, and that they have a bias towards the upper source for small ISDs.

There are two kinds of undershoots:

1. *One observed for single source where subjective lateral position undershoots true position. This can be reduced by freeing cat’s head.*
2. *With two sources, there is an additional undershoot in which the subjective lateral position is pulled towards the lag side rlative to the subjective position of leading source. This is also seen in human.*

---

[4] Summming localization is the perception of a “phantom” sound located between the sources. (Tollin *et al*, 2003)

[5] Localization dominance is the phenomenon where the subject somehow restricts the computation of location to that based on the spatial cues of the sound that arrives at the ears first from the true source. (Tollin *et al,* 2003)

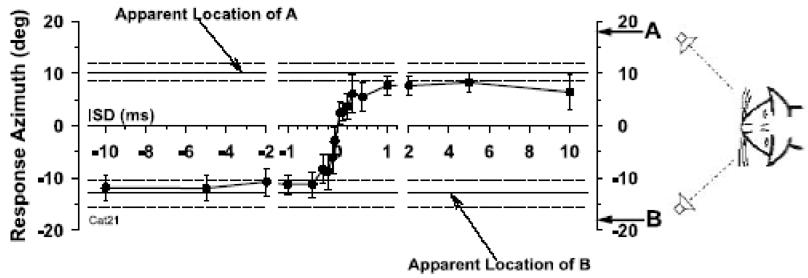

Figure 5 Mean final eye positions for paired sources in the azimuths as a function of ISD. (Tollin *et al*, 2003, Figure 2C)

The authors also tried to ascertain the echo threshold[6] and they found that for ISDs about 15ms and greater, the cats often made saccades to the lagging source on some trials and toward the leading source on other trials. In Figure 6, we see that the bimodal distributions of responses effectively pulled the mean response azimuths toward the midline at around 15ms. The authors also argue that their method would likely overestimates echo threshold because the cats may have been reluctant to saccade toward the lagging source location if *it produced a weak percept*. These data reveal that cats experience the PE phenomena similar to humans.

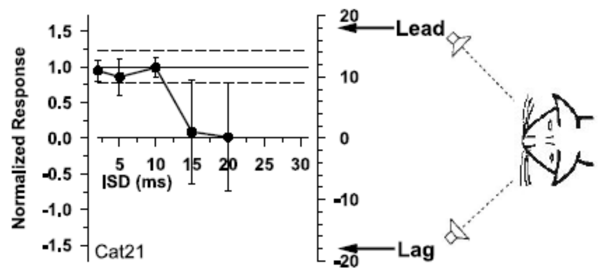

Figure 6 Mean response azimuth toward leading source location for paired-source stimuli positioned along horizontal plane as function of ISD. (Tollin et al, 2003, Figure 8)

With the three major perceptual phenomena of the PE illusion observed in cats, it is hoped that the ISD ranges found could be used to interpret some physiological studies on the PE. Fitzpatrick *et al* (1999) presented neural responses to paired clicks with varying conditioner-probe intervals (CPI, c.f. ISD) from several structures of the ascending auditory system in unanesthetized animals to compare with the behavioral phenomenon of PE. The authors found that there is lengthening of *the average* recovery times as one ascends the auditory pathway (see Figure 7). 50% recovery time, defined as the response to the probe relative to that of an identical stimulus presented in isolation, was used as a benchmark to compare the different single unit neural recordings from different levels. *[not exclusively – They used mainly average recovery curves]* Recovery functions also vary in shape. At the auditory nerve (AN) level, it is homogeneously monotonic (Figure 8 Left), and as it ascends to higher pathway, the recovery functions become more heterogeneous, such as that in Figure 8 (Right), found in the inferior colliculus (IC), compared to that of the AN level.

[6] Echo threshold is the shortest delay at which a reflection is localized near its source. (Tollin *et al*, 2003)

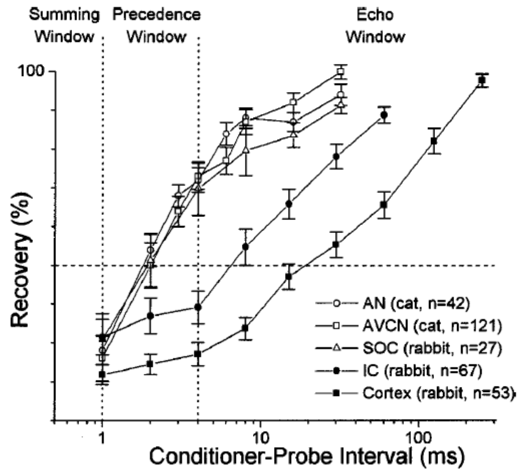

Figure 7 Population recovery functions for each level compared. (Fitzpatrick *et al*, 1999, Figure 3)

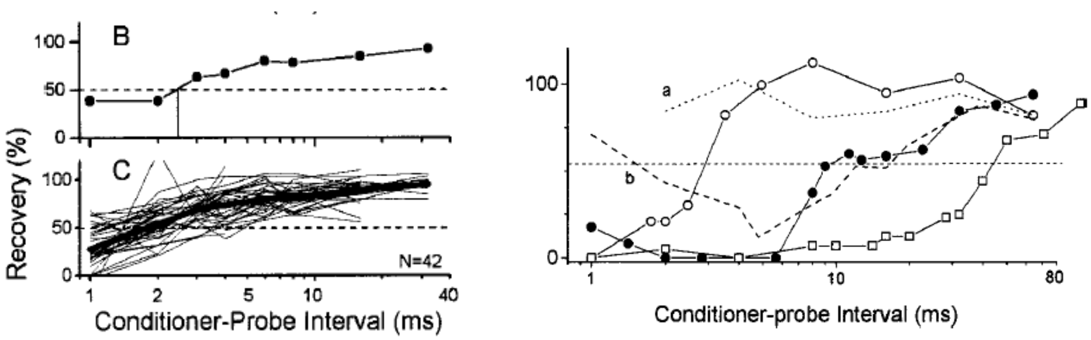

Figure 8 (Left) Recovery function for one auditory-nerve fiber (top), recovery function for all nerve fibers recorded (bottom). (Right) Examples of recovery functions seen in the IC. (Fitzpatrick *et al*, 1999, Figure 2 B, C and Figure 5)

There are some experimental uncertainties in the Fitzpatrick *et al* (1999) paper that make the interpretation of some of their findings ambiguous. In their study, the AN and the anteroventral cochlear nucleus (AVCN) were studied in decerebrate cats while the superior olivary complex (SOC), IC and auditory cortex were studied in *awake* rabbits. It is argued that both preparations were unanesthetized, which is an important point of similarity. However, the species differences make the findings hard to interpret. At that time, there were no cat behavioral studies *[there were some, but they were weak]*, and even now, there are still no rabbit behavioral studies on PE. The authors postulate that the PE relies on the ability to detect activity in the most sensitive neurons in higher structures of the auditory pathway. However, without the corresponding behavioral animal models, such hypothesis would be hard to verify. *[I see no reason why rabbit should be fundamentally different from cats and humans with respect to PE].*

The Jeffress coincidence-detector model[7] is 'currently the backbone of virtually all computational models of sound localization' (Colburn, 1996 in Joris *et al*, 1998). The existence of such an arrangement has been confirmed in the avian nucleus laminaris (NL), but there are conflicting evidence that its mammalian counterpart, the medial superior olive (MSO), relies more on the glycinergic inputs from the medial, and partially from the lateral, nucleus of the trapezoid body

[7] This model was based on three main assumptions: 'bilateral, time-locked or phase-locked inputs into the ITD-processing system; coincidence detection by ITD detector neurons; and an arrangement of delay lines to adjust coincidence detector neurons to different preferred ITDs, creating a tonographic representation of azimuthal space.' (Grothe, 2003).

(MNTB, LNTB) (Brand *et al*, 2002) to improve precise ITD coding *[not to improve coding but to create internal delays]*. In the Jeffress model, it is generally assumed that response maxima of ITD functions are independent of frequency-tuning characteristics of the neurons. However, from the MSO recordings performed on the Mongolian gerbil, a rodent with well-developed low-frequency sound localization, Brand *et al* (2002) found that the ITD response function of most neurons peaks outside the physiologically relevant range (Figure 9a). Furthermore, the ITD that evokes the peak response is dependent on neuronal tuning for sound frequency (Figure 9b).

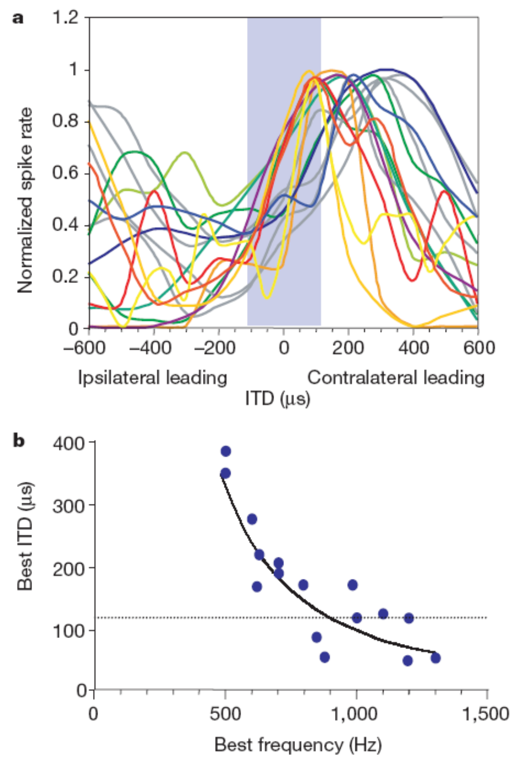

Figure 9 Peaks of ITD functions are outside of the physiologically relevant range. (a) ITD functions of gerbil MSO neurons, each tested at its best frequency. (b) Distribution of best ITDs as a function of best frequencies of the neurons. (Brand *et al*, 2002, Figure 2)

From Figure 9, it is also evident that the maximum dynamic range for each neuron's ITD-evoked response is within the physiologically relevant range. Using local application of strychnine, a glycinergic antagonist, by iontophoresis, Brand *et al* (2002) found that strychnine produced a significant increase in discharge rate, and that the peak of the ITD functions were shifted towards zero, and thus the steepest slope of the function fell outside the physiologically relevant range of ITDs (See Figure 10). In this case, the dynamic range within the physiological range was reduced from 83% to 21%. The authors argued that the glycinergic inputs to the MSO must be precisely phase-locked to the excitatory inputs, but the contralateral inhibition has also been shown to be more prominent than the ipsilateral inhibition. Using a modified Hodgkin-Huxley point neuron model based on physiological findings in neurons of the AVCN, the authors found that by increasing the strength of the timed inhibition, the ITD sensitivity maximal point can be shifted towards longer values of ITD. The excitatory and the inhibitory synaptic time constants were both set at 0.1 ms, which is similar to the values found by Kalluri *et al* (2003) in their onset-neuron simulations. The authors concluded from their data that ITD tuning in the gerbil MSO is the product of an interaction of precisely timed excitatory and inhibitory inputs. *[0.1 ms is a very short time constant for an inhibitory input, much shorter than observed* in vivo*].*

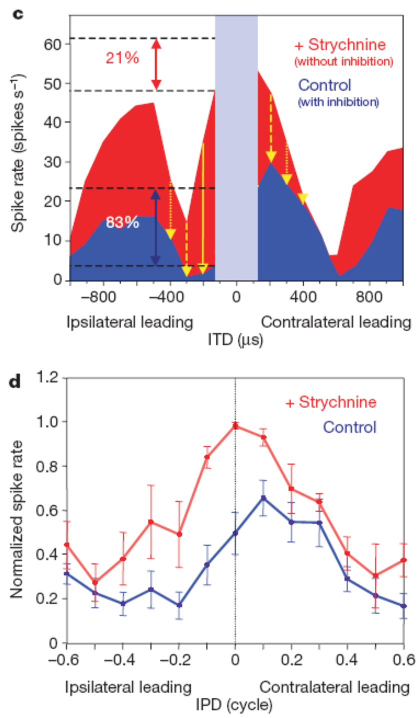

Figure 10 (Top) ITD functions of a neuron during control condition and strychnine application. (Bottom) Averaged interaural delay functions under control conditions and during application of strychnine. (Brand *et al*, 2002, Figure 3c, d)

Other points learnt from Theme 3:

- Behavioral studies on psychoacoustic phenomena, like the PE, on non-human subjects can be useful for comparing physiological data obtained from the same non-human species. However, if a physiological study is performed on different animals, the results can be hard to interpret. *[The results of Fitzpatrick et al would be hard to interpret even if all data were from the cat – the problem is that the relationship of neural echo suppression to the PE is unclear].*
- Derived values, such as cue weights (Macpherson *et al*, 2002), need to be thought out carefully. In this example, the cue weight closed to, but smaller than, unity can be due to either the effect of the other binaural cues or the effect of undershooting, observed in other psychophysical studies.
- *Undershooting is small in humans. Also it could be taken into account in computing cue weights by using subjective position as reference.*

Papers discussed:

Bernstein LR (2001). Auditory processing of interaural timing information: New insights. *J Neurosci Res* **66**:1035–1046.

Brand A, Behrend O, Marquardt T, McAlpine D, and Grothe B (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature* **417**:543-547.

Fitzpatrick DC, Kuwada S, Kim DO, Parham K, and Batra R (1999). Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex. *J Acoust Soc Am* **106**:3460-3472.

MacPherson EA and Middlebrooks JC (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J Acoust Soc Am* **111**:2219-2236.

Tollin DJ and Yin TCT (2003). Psychophysical investigation of an auditory spatial illusion in cats: The precedence effect. *J Neurophysiol* **90**: 2149–2162.

Reference:

Ebata M., (2003). Spatial unmasking and attention related to the cocktail party problem. *Acoust Sci. & Tech.* **24** (5): 208-219.

Freyman R., Helfer K., McCall D., Clifton R., (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.* **106** (6): 3578-3588.

Grothe B (2003). New roles for synaptic inhibition in sound localization. *Nature Rev Neurosci*. **4**:540-50.

Joris P., Smith P., Yin T., (1998). Coincidence Detection in the Auditory System: 50 Years after Jeffress, *Neuron* **21**:1235-1238.

Kalluri S and Delgutte B (2003). Mathematical model of cochlear nucleus onset neurons: I. Point neuron with many, weak synaptic inputs. *J. Comput. Neurosci*. **14**:71-90.

Litovsky R., Shinn-Cunningham B., (2001). Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression. *J Acoust Soc Am* **109**:346-358.

Reyes AD, Rubel EW and Spain WJ (1996). In vitro analysis of optimal stimuli for phase locking and time-delayed modulation of firing in avian nucleus laminaris neurons. *J Neurosci* **16**:993-1007. (Discussed in class, not included in theme paper)