Original Article

H. Steven Colburn
Barbara Shinn-Cunningham
Gerald Kidd, Jr.
Nat Durlach

Hearing Research Center, Boston University, Boston, MA, USA

The perceptual consequences of binaural hearing

Las consecuencias perceptuales de la audición binaural

Key Words

Binaural hearing
Interaural time difference
Interaural level difference
Complex acoustic environments
Informational masking

Abbreviations

ITD: Interaural time difference
IPD: Interaural phase difference
ILD: Interaural level difference
JND: Just-noticeable difference
MLD: Masking level difference
CN: Cochlear nucleus
SOC: Superior olive complex
SRT: Speech reception threshold
CRM: Coordinate response
   measure
HI: Hearing impaired
NH: Normal =hearing

Abstract

Binaural processing in normal hearing activities is based on the ability of listeners to use the information provided by the differences between the signals at the two ears. The most prominent differences are the interaural time difference and the interaural level difference, both of which depend on frequency. This paper describes the stages by which these differences are estimated by the physiological structures of the auditory system, summarizes the sensitivity of the human listener to these differences, and reviews the nature of the interaural differences in realistic environments.

Sumario

El procesamiento binaural durante actividades auditivas habituales está basado en la habilidad de los oyentes para usar la información proporcionada por las diferencias entre las señales en ambos oídos. Las diferencias más prominentes son las de tiempo y de intensidad interaural, ambas dependientes de la frecuencia. Este trabajo describe las etapas por las cuales estas diferencias son estimadas por las estructuras fisiológicas del sistema auditivo; resume la sensibilidad del oyente humano a éstas diferencias y revisa la naturaleza de las diferencias interaurales en ambientes reales.

The fact that the signals at the two ears are different leads to a number of advantages in listening binaurally. Some of these advantages are related to the simple ability to exploit the location of the better placed ear. Other advantages are based on the ability to extract information from the differences between the signals. This extraction involves the analysis of the interaural differences in the waveforms, i.e., binaural processing, which is the focus of this paper. These differences are usually described in terms of the interaural time difference (ITD), the interaural level difference (ILD), the interaural cross-correlation coefficient (ICC), and the energy in the difference in the waveforms. Each of these quantities are generally considered as a function of frequency, specifically computed from bandpass filtered versions of the input stimuli. These parameters are not independent, many are related to the interaural cross-correlation function (ICF), and they are frequently used to understand, characterize, and model binaural phenomena.

The initial processing of auditory inputs is summarized briefly here in terms of signal processing operations. Acoustic inputs are filtered to generate to a set of narrowband waveforms that are transduced to generate neural firing patterns on the auditory nerve. These patterns are further processed in the brainstem nuclei, resulting in neural patterns that are sensitive to the interaural differences in the narrowband filtered stimuli. Thus, the result of the peripheral and brainstem processing is a neural representation that is equivalent to a sequence of estimates of interaural time and level differences for each frequency band. This is the basic input to the binaural processor.

These initial stages of binaural processing are relatively well understood and provide the input for central processing that interprets the interaural-difference information to generate perceptions that facilitate interpretation of the acoustic world. The interpretation of this information involves a combination of bottom-up and top-down processing. In order to understand more about this processing, it is useful to start with simple environments containing single sources with minimal reverberation and then address the more complex environments that involve multiple, simultaneous sources and significant reverberation. Real environments, which generally include multiple sources and reflections, are very complicated acoustically and raise challenges for listeners, for scientists attempting to understand and model hearing in such environments, and for

H. Steven Colburn
Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215.
E-mail colburn@bu.edu

engineers attempting to develop artificial processing systems to assist listeners in such environments. Although these challenges are substantial even when the listener has normal hearing, they often become crucial when the listener suffers from a significant hearing impairment.

Basic binaural physiology

Peripheral narrowband filtering

The cochlea filters sound into relatively narrow (one-sixth to one-third octave) frequency bands, creating a parallel representation of the input broken down into mutiple frequency channels. The resulting frequency-based representation of sound affects all subsequent stages of neural auditory processing and auditory perception, including binaural processing and binaural hearing. As a result, understanding binaural hearing depends upon understanding the nature of narrowband signals. The effective bandwidth of auditory peripheral filtering increases roughly proportionally with the center frequency of the channel, so that low-frequency channels have a narrow bandwidth (in Hz) compared to high-frequency channels.

The impact of peripheral bandpass filtering of acoustic stimuli is profound, particularly because filters are relatively narrow so that waveforms have well-defined envelopes. Basically, the acoustic input is filtered to give narrowband stimuli with several important properties. First, the fine structure (the rapid oscillations at the center frequency of the narrowband filter) and the envelopes of the signal (corresponding to the time-varying amplitude of the oscillations) can be specified almost separately as two distinct time functions. Second, the phase of the fine structure varies with time. This phase variation corresponds to small variations in the length of individual cycles of the fine structure in the waveforms (i.e., in the deviations around the average cycle length). Although this phase variation is a general property of narrowband waveforms, it is less evident than the envelope and the fine structure oscillations. Third, because the signals are narrowband, the envelopes and the phase-variations are relatively smooth functions of time.

The maximum rates of variation in both the envelope and phase increase with the bandwidth of the signal. When a broadband signal is presented to a listener, the effective bandwidth of the signals represented in each frequency channel is determined solely by the bandwidth of the peripheral filter, so the maximal rates of envelope and phase variation depend on the center frequency of the channel being considered. In particular, because the bandwidth (in Hz) is broader for high-frequency channels compared to low-frequency channels, the variations in the envelope and phase are generally more rapid in the high-frequency channels.

If the input is broadband relative to the auditory peripheral filters, the bandwidths of the filtered signals are determined by the auditory periphery. However, if the input signal has a narrow bandwidth relative to the auditory peripheral filter responsive to that input, the input signal bandwidth will determine the bandwidth of the signal in the auditory periphery, and the envelope variaion will be slower. For instance, for tonal stimuli, the envelope and the phase in the peripheral representation are constant (since the stimulus bandwidth is zero) except for transients at the beginning and end of the stimulus. Similarly, for purely amplitude-modulated waveforms, there is no phase variation in the peripheral responses of the auditory system, and for purely frequency- or phase-modulated waveforms, there is no envelope variation.

The firing patterns of the primary auditory nerve, i.e., the eighth or cochlear nerve, can be modeled by rectifying and low-pass filtering the narrowband filter outputs and using the result as the rate of a random firing generator. At all frequencies, the firing rate of the auditory nerve varies with the envelope (i.e., essentially with the short-term amplitude of the signal). The exact timing of individual spikes depends on the center frequency of the corresponding filter. For low center frequencies, the neural responses are highly synchronized to the fine structure. In contrast, at high center frequencies, neural responses do not follow the fine structure because of temporal limitations in the biophysical properties of cells in the auditory periphery. Instead, in high frequency channels, the sometimes rapid fluctuations in the signal envelope cause synchronous bursts of neural spikes that track the envelope shape. The cut-off frequency at which the fine-structure timing information begins to decline varies with species, and is near 800 Hz in the cat (although some synchronization to the fine structure is seen above 4 or 5 kHz) and 8 kHz in the barn owl. For human listeners, the lack of perceptual sensitivity to interaural phase (and to any ongoing fine structure delays) above about 1.5 kHz is often modeled as a consequence of the loss of synchronization as frequency increases. The fact that listeners can discriminate ongoing time delay for narrowband high-frequency waveforms is consistent with the use of synchronization to the envelopes of high-frequency waveforms.

Band-by-band processing

Both physiological and perceptual empirical results support the idea that binaural comparisons are made only between frequency-matched, narrowband inputs from the left and right ears. This postulate, that interaural differences are computed from comparisons of inputs from the same frequency bands in each ear, is supported by both psychophysical and physiological data. Specifically, when narrowband waveforms with different center frequencies are used in interaural time difference (ITD) sensitivity experiments, performance rapidly degrades if the center frequencies of the inputs become separated by more than the commonly assumed bandwidths of the peripheral auditory filters (c.f., Nuetzel and Hafter, 1981). Physiological experiments show that binaurally sensitive neurons are driven by ipsilateral and contralateral inputs that generally have the same frequency tuning (Guinan et al, 1972; Goldberg and Brown, 1969; Boudreau and Tsuchitani, 1970; Yin and Chan, 1990), again consistent with frequency-band-by-frequency-band binaural comparisons.

Given such results, it is reasonable that research focuses on interaural comparisons of frequency-matched (or nearly matched) narrowband waveforms. If the waveforms have the same center frequency, then the differences in the waveforms can be described by the differences in the ratios of the envelope functions (to which listeners are sensitive at all frequencies) and the phases of the fine structures at low frequencies (where listeners are sensitive to phase). The differences in the envelope intensities correspond to Interaural Level Differences (ILDs) and the differences in the fine structure phases correspond to Interaural Phase Differences (IPDs). At higher frequencies, the

IPDs are not available (due to loss of fine-structure sensitivity); however, listeners are sensitive to Interaural Time Differences (ITDs) in the envelopes at high-frequency channels. These quantities are generally functions of time, although they vary slowly for narrowband waveforms.

Brainstem physiology

This section briefly summarizes attributes of neurons in the ascending auditory pathway that are known to have a significant effect on binaural processing.

As described above, auditory peripheral filtering transforms an input stimulus into a frequency-based representation in parallel channels. Moreover, this representation is inherently stochastic. Thus, even while the expected temporal patterns of neural firings generally change with a change in the stimulus waveform, the change in the stimulus may or may not be detectable to an organism because of the stochastic variability of the neural response. In the cochlear nucleus (CN), an obligatory center in the ascending auditory pathway, evidence (e.g., Joris et al, 1994) shows that the timing of many individual neural responses is more precise than the timing of the primary neurons innervating the CN.

Activity from the CN is further processed to extract binaural information in the medial superior olive (MSO). MSO neurons are sensitive to IPDs of low-frequency input stimuli (providing a neural mechanism for encoding IPDs and ITDs), while neurons of the lateral superior olive (LSO) are sensitive to the ILDs of input stimuli (and to some temporal delays, as described below).

The average firing rate of a typical MSO neuron is maximized when the input stimuli have a particular ITD, known as the neuron's "best ITD." Because the inputs driving MSO neurons are narrowband signals, the neuron also typically responds well to stimuli whose ITD is equal to the best ITD plus or minus multiples of a full cycle of the center frequency of the inputs and has a minimum response when the input signals equal the best ITD plus (or minus) one half of a cycle (e.g., as if, in effect, the neuron is sensitive to IPD rather than ITD). MSO sensitivity to ITD is thought to arise through a coincidence mechanism, an idea originally suggested by Lloyd Jeffress in the 1940's (Jeffress, 1948). The basic structure suggested by Jeffress is consistent with much of the available anatomy and physiology of the MSO (Joris et al, 1998). This mechanism could also be used to generate sensitivity to time delays in envelope waveforms in high frequencies; although, the majority of neurons in the MSO are tuned to low frequencies (Guinan et al, 1972). Current issues in the modeling of these phenomena include the distribution of best ITDs and the role of inhibition (Harper and McAlpine, 2004; Grothe, 2003).

Neurons in the LSO are generally inhibited by contralateral inputs and excited by ipsilateral inputs (and are therefore known as IE cells). As a result, the population of LSO neurons provides a natural structure for extracting information about ILDs. In particular, if an input sound is more intense in the ipsilateral ear than the contralateral ear, then the excitatory input will be strong relative to the inhibitory input, and the neuron will fire strongly. If the relative intensities of the sound at the two ears is made more equal, the inhibition to the LSO will grow, and the neuron's response will decrease. Thus, if the activity of the left LSO is greater than the activity of the right LSO, the stimulus is likely to be more intense in the left ear than the right ear.

Although the LSO may be very small in humans (Moore et al, 1999), this same binaural-opponent-processing scheme could be realized in neurons at higher levels of the ascending human auditory pathway (e.g., at the level of the inferior colliculus (IC), which receives strong contralateral inputs from the LSO in lower mammals). LSO neurons are also sensitive to onset ITD as well as ILD, and a neural network of such cells provides sensitivity to combinations of ITD and ILD. Since onset and offset ITDs involve time-varying ILDs, sensitivity to ILD implies some sensitivity to onset/offset ITD and to ITDs in some types of amplitude-modulated waveforms.

Olivary neurons project to the inferior colliculus (IC), like all neurons in the ascending auditory pathway, and neurons in the IC show a range of binaurally sensitive responses. Because the IC is relatively easy to access for physiological recordings, there is much more information available about the activity of neurons in this nucleus that for other populations of binaurally sensitive neurons in the brainstem. Also, the nuclei that we have already discussed send well-defined inputs to the IC and it is presumed that binaural information about IPD/ITD and ILD converges and is integrated in the IC (consistent with evidence from avian IC, where cross-frequency integration of binaural information is well documented).

Basic binaural abilities

Sensitivity to fixed interaural differences

Both historically and logically, the binaural parameters of primary interest are ITDs and ILDs. Therefore, characterizing the Just-Noticeable Differences (JNDs) in the ITD (or IPD) and in the ILD is an important step in quantifying binaural abilities. For low-frequency stimuli with durations of tenths of seconds, results from many laboratories show that the best listeners typically achieve ITD JNDs of roughly 15 µs and ILD JNDS of about 0.5 dB (c.f., Durlach and Colburn, 1978) however, as discussed below there are many subjects who do not achieve these levels of performance (e.g., Bernstein et al, 1998).

The 15-µs ITD JND is achieved for almost all stimuli that include low-frequency energy: noise bands, clicks, and tone bursts with frequencies between 500 Hz and 1 kHz. For pure tones, the ITD JND increases with increasing frequency and is not measurable for tone frequencies above about 1.5 kHz. JNDs in IPD are generally predicted by converting the ITD JND into phase. The loss of IPD sensitivity with increasing center frequency is consistent with the loss of synchronization to fine structure as frequency increases. For broader bandwidth stimuli, sensitivity to envelope ITD is supported by the neural synchronization to the stimulus envelope. The envelope ITD JND depends critically on the shape of the envelope (e.g., Bernstein), but is generally somewhat greater than the low-frequency ITD-JNDs (e.g., on the order of 100 µs for the best listeners (Henning, 1974; Koehnke et al, 1986).

In contrast to ITD JNDs, ILD JNDs are approximately independent of frequency, equal to roughly 0.5 dB at all frequencies. As long as the stimuli are audible and at a comfortable listening level, effects of level on both ITD and ILD sensitivity are relatively weak.

The most obvious effect of changing ITD and/or ILD is to change the perceived lateral (left-right) position of an auditory image, which is heard nearer to the ear with the leading or more intense stimulus. One can "trade" the effects of ITD and ILD to

create images in the center of the head (e.g., with a more intense, lagging sound in the right ear). However, when unnatural combinations of ITD and ILD are presented, the resulting sound may be reported as being at one location but will have other attributes that differentiate it from a sound with consistent spatial cues. For instance, in addition to the dominant lateral location, listeners often hear the source image as wider, with a secondary lateral location, with a complex asymmetric image with "tails" toward one ear, etc. These results show that listeners do not only have access to "the" lateral location of a sound source, but also perceive other information (e.g., the distribution of ITD and ILD cues present in a stimulus).

Range of "normal" binaural abilities
Although normal-hearing, practiced, "good" listeners all have similar low JNDs, performance levels vary widely in a more general population, perhaps reaching hundreds of microseconds (for ITD JNDs) and several decibels (for ILD JNDs) (c.f., Bernstein et al, 1998). When testing a clinical population of listeners with hearing impairments, it would be expected that the same range of abilities is likely to be observed independent of any impairments so that some hearing-impaired listeners may have large binaural JNDs that are unrelated to their hearing loss. Not only is it unclear as to what level of performance is "normal," there is little predictability about the effects of hearing loss on binaural hearing. Some listeners with severe sensorineural hearing loss (80 dB HL) have normal ITD JNDs, while some listeners with moderate losses have high ITD JNDs (Hausler et al, 1983).

Sensitivity to temporal variations in interaural differences
In addition to the sensitivity to fixed values of ITD and ILD described above, listeners are also sensitive to temporal fluctuations in the ITD and ILD values. The rate of variability in ITDs and ILDs are constrained by the bandwidth of a signal, and the detectability of the fluctuations are dependent on both their bandwidth and amplitude (the amplitude of the fluctuations is related to the standard deviation).

When a normal-hearing listener is presented with a low-frequency stimulus having fluctuating interaural differences, the perceptual impression depends on the rate of the fluctuations. For slow fluctuations, one perceives a moving image as one would expect from the lateral impressions created by fixed interaural differences. When the fluctuation rate increases above a few Hertz, listeners have difficulty following the trajectory of the sound and report a broadening of the sound image. In this rapid fluctuation domain, the temporal variations of ITDs or ILDs translate to a width impression. The fact that the limit of tracking the lateral displacement is so low, is referred to as "binaural sluggishness." Note that, although the lateral displacement behaves "sluggishly," the fluctuations in the interaural differences are detected and have an effect on perception; thus, the binaural system is not insensitive to these rapid fluctuations. For stimuli with randomly varying interaural differences, the stimulus can be characterized in terms of the interaural correlation, as discussed in the next section.

INTERAURAL CORRELATION DISCRIMINATION
Measurements of sensitivity to interaural correlation have also been used to characterize binaural capabilities. These measurements typically use stimuli that are created by combining statistically independent noises to create stimuli with varying amounts of common components. The correlation coefficient is equal to the proportion of interaurally common components (for positive values) so that identical noise has unity correlation and statistically independent noises have zero correlation.

Listeners are very sensitive to decreases in interaural correlation and can detect the change from a value of unity (1.00) to 0.96 (Pollack and Trittipoe, 1959; Gabriel and Colburn, 1981; Goupell and Hartmann, 2006). The perception of a correlation decrease is usually an increase in the perceived width of the auditory image.

The abilities of listeners to detect changes in the distribution of the temporal sequence is apparently the basis of the discrimination of interaural correlation. The analysis of interaural discrimination in terms of the distribution of interaural time and intensity differences can be used to analyze the performance in terms of sensitivity to the width of the distribution of interaural differences. (Zurek, 1991; Goupell and Hartmann, 2006).

BINAURAL DETECTION (MASKING LEVEL DIFFERENCES)
Binaural detection experiments are a classical measure of binaural hearing and illustrate some of the advantages of binaural listening in avoiding interfering effects of a masking noise. When the target and masker have interaural parameters that are different, there can be large advantages of detection using two ears. These advantages are described as Masking Level Differences (MLDs), which are defined as the differences in thresholds (in decibels) for cases in which the ears receive identical signals and thresholds for the binaural cases with different interaural relationships for target and noise. One of the popular reference conditions, which shows the largest improvement in the binaural case, is the "antiphasic case," which corresponds to an interaurally identical noise waveform with an interaurally out-of-phase target. This case, and others like it, show binaural improvements that are much larger than the improvements afforded by the better-placed ear. With narrowband noises, the advantages can be as large as 20 decibels.

Binaural detection performance and interaural correlation sensitivity can be related for many cases (Durlach et al, 1986). For example, the decorrelation within the target band of frequency in binaural detection is about the same as the decorrelation within a target band that is statistically decorrelated. Quantitative analysis shows that performance levels in many different situations are consistent in the sense that the correlation change sensitivity and binaural detection thresholds are consistent, considering the description of interaural cross-correlation sensitivity in terms of fluctuations in interaural difference parameters. In other words, one way to describe the cue for binaural detection abilities is in terms of the increase in the variability of the interaural time and level differences (ITDs and ILDs).

Of course this means that the binaural detection experiments can also be directly related to the distributions of the ITDs and ILDs. When there is no target present and the stimulus is the masker alone, the interaural differences are fixed at the values corresponding to those of the masker. When the target is added, the interaural differences vary randomly in time according to the random phase angle of the masker waveform relative to the

target tone. This relationship can be seen in Figure 1, where the temporal distributions of ITD values are shown for the tone-in-noise detection case (from Shinn-Cunningham, 2005). The upper panel shows that all ITDs are near zero when target and masker are both interaurally correlated. The lower panel shows that the addition of the tone from roughly 20 to 80 ms causes significant variability in the value of the ITD during this interval.

Binaural hearing in simple environments

Single sources in anechoic space

The relation between interaural time and level differences and the physical environment is relatively straightforward for single sources in anechoic environments. In these circumstances, the interaural time and level differences can be plotted as simple functions of frequency for each spatial location of the source. When the source is farther away than about a meter, these functions are independent of distance and are easily character-ized. When the source is closer, within about one meter from the head, the dependence on distance is substantial and requires more detailed characterization (c.f., Shinn-Cunningham et al, 2000). To a first approximation, one can use a spherical model of the head to predict the interaural differences; although, the interaural differences vary significantly from individual to individual, depending on the overall size of the head as well as on the detailed shape of the head and pinnae.

The dependence of the ITD on frequency for distant sources is small enough that it is often ignored. To a first approximation, the ITD for frequencies above about 2 kHz is about what one would compute from the differences in pathlength from the source to the ears, and the ITD increases relatively smoothly as frequency decreases to reach a low-frequency asymptote about 50% higher than the high frequency value. This general behavior applies to all locations.

The dependence of the ILD on frequency is more variable, even for distant sources, and the shape is more variable from location to location. In general, the ILDs show larger variation when the frequency is higher as one would expect. When a spherical head model is used to approximate ITD and ILD, the

simplest case assumes that the ears are at opposite ends of a diagonal of the sphere. In this case, the physical system is symmetric about the interaural axis and contours of constant ILD or ITD correspond to cones for distant sources. This concept is generalized for actual heads to describe the contours of constant interaural differences, which are still called "cones" of confusion even though the contours may be complex shapes

For sources near the head, the interaural differences are approximately constant on donuts (or, more formally, tori) of confusion since the interaural differences vary with distance as well as with angle from the interaural axis, even for a spherical head. The dependence of interaural differences in this case is more complex to describe. This topic is treated more completely by Shinn-Cunningham and colleagues (Shinn-Cunningham et al, 2000).

For simple acoustical environments (one dominant source and anechoic space), many of the perceptual consequences of binaural processing can be understood by consideration of the binaural information contained within the stimulus waveforms and as extracted by the auditory system. In contrast to this relatively well understood situation, the situations with multiple sources or in reverberant environments are not well understood.

Weak target and strong masker in anechoic space

In the usual noise-masked detection experiment, the signal-to-noise ratio is usually low, so that the signal can be described as a perturbation to the acoustic field that is dominated by the masker waveform. In this case, the acoustic space is still relatively simple. Specifically, the mean values of the interaural differences are determined by the masker waveform according to its position relative to the head. The presence of the weak target causes relatively small changes in the interaural differences. These changes may be shifts to constant values near those of the masker alone, which description applies to situations where the target and masker are perfectly correlated and synchronized. Alternatively, for cases like a tonal masker in a random noise band, the changes may be fluctuations around the stable values of the masker, as we described and illustrated above (c.f., Figure 1).

Complex acoustic environments

Listening environments that are more complex than those considered in Sec. 4 can arise as a consequence of: (1) an increase in the number of masking sources, (2) the acoustic environment being reverberant rather than anechoic, and/or (3) the sounds emanating from the various sources become more complicated. In this section, we consider some phenomena associated with each of these factors and discuss their relationship to binaural hearing.

While the focus continues to be on interaural differences, the complexity of the ITDs and ILDs encountered in these environments is often too great to be usefully analyzed. The problems arise from the fact that the interaural differences combine in complex ways even though the waveform pressures simply add. In the simplest case of two fixed sinusoids of the same frequency coming from different directions, the Interaural Differences (IDs) are constant and they can be calculated using simple trigonometry. The resulting IDs depend on the relative levels and phases of the tones at the two ears. If one signal is significantly more intense than the other, then the resultant IDs are, as
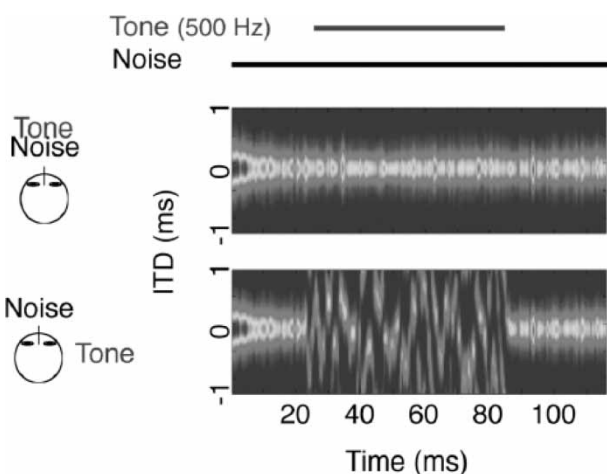

Figure 1. Temporal distributions of ITD values for tone-in-noise detection case for 500-Hz center frequency. Figure from Shinn-Cunningham, 2005.

expected, approximately the same as the interaural differences of the stronger signal by itself (as determined by the location of the dominant source relative to the head). If the signal levels are comparable at a single ear, the combined signal can have almost any amplitude from the sum of the amplitudes to the difference. This leads to an extremely large range of possible interaural level differences. Similarly, the resultant interaural phase difference can be outside the range of the interaural phases of the individual source waveforms. When the two sources are variable, or are fixed sources with random noise waveforms, the situation is even more complex, and the interaural differences vary randomly. A similarly complex pattern of interaural differences are generated in reverberant rooms, as discussed further below.

Multiple maskers
In the discussion of simple environments, the only circumstance considered in which there was more than one sound source was the case in which the perception of a single target source was degraded by the presence of a single, much stronger masking source. In the preceding discussion, we considered the case of two nearly identical sources. From both practical and theoretical viewpoints, it is important to consider conditions in which more than two sources are present and, in particular, in which the attempt to listen to the target source is impeded by the presence of two or more masking sources.

In general, the question of how many masking sources are present, i.e., how the total masking signal can most reasonably be decomposed into a sum of individual maskers, is a complicated one because it involves the notion of an auditory "object" or "stream" (discussed further below). Here, we sidestep this question by simply assuming that there are N independent maskers, each of which evidences a distinct set of interaural relations (ITD and ILD).

These cases are difficult to analyze but are of considerable practical significance, and increasing attention is being directed toward both empirical and theoretical understanding of interference from multiple maskers. There are data from lateralization (Langendijk et al, 2001), from speech intelligibility (Peissig and Kollmeier, 1997; Hawley et al, 1999; Freyman et al, 2004) and from simple detection. In general, these data demonstrate that the perception of a target is greatly degraded as the number N of maskers increases from N = 1 to N = 2 (provided, of course, that the difference in interaural relations for the two maskers is significant). In view of the fact that it is theoretically possible to cancel out N-1 maskers (but no more than N-1) with N spatially separated sensors, and the human listener has N = 2 ears, this general result is not at all surprising. Specific analyses of these cases is difficult, however, in part because of the complexity of the interactions of the stimulus components. As noted above, the interaural differences combine in non-linear ways and the spectro-temporal pattern of the interaural differences is very complex. The fact that many normal environments, including classrooms and restaurants, contain multiple sources is an incentive for more specific consideration of listeners' abilities in these environments.

Reverberation
Because most environmental spaces have significant reverberation, it is useful to consider the effects of reverberation on binaural processing in general and on interaural differences in particular. Consider first the case when there is a single noise source at three alternative directions relative to the head. The ITDs are evaluated in a room for several levels of reverberation using software simulations. Figure 2 (taken from Shinn-Cunningham and Kawakyu, 2003) shows the distributions of short-term ITD values as detected by a neural circuit model for three directions and three conditions of reverberation. The neural
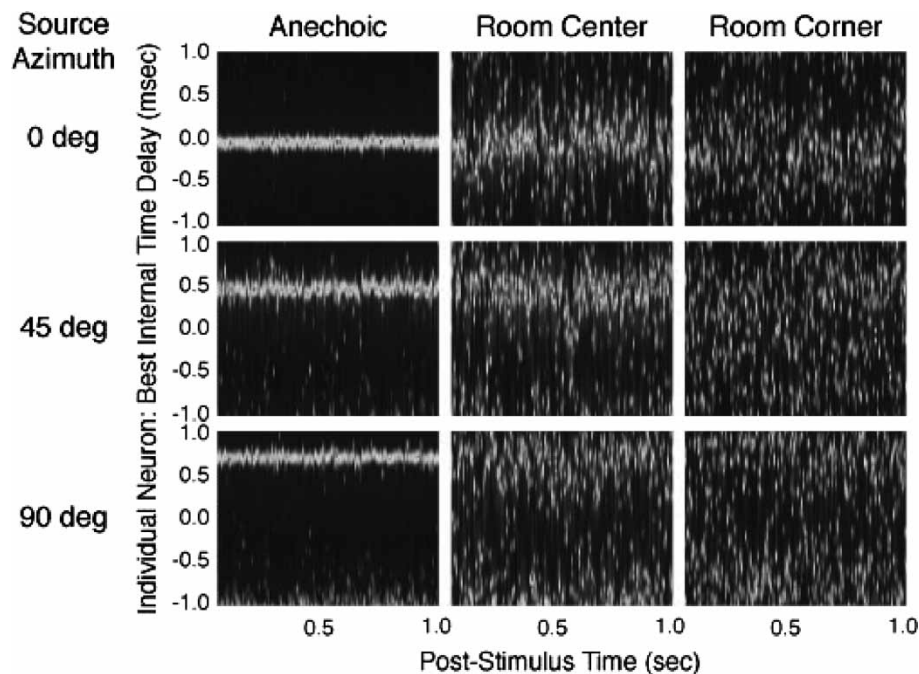

Figure 2. Temporal distributions of ITD — measured for noise in a room From Shinn-Cunningham & Kawakyu, 2003.

circuit is tuned to the 547-Hz frequency band and the processing is based on short-time estimates (four cycles of the center frequency). When there is minimal reverberation, as shown in the left column (anechoic), the ITD values are all relatively close to the ITD value corresponding to the direction of the source. The network response shows that in this case the ITDs are all near zero for the 0-degree (straight ahead) direction, near 0.5 ms for the 45-degree direction, and near 0.7 ms for the 90-degree direction. The middle column shows the effects of reverberation for a location near the middle of the classroom that was used for this simulation. It should be apparent that the distribution of ITD values over time show a relatively strong dispersion throughout the approximately one-second stimulus waveform. The values are concentrated near the anechoic values but there is substantial variability. The right column shows the effects of stronger reverberation, as calculated for a location near the corner of the room. In this case, the variability of the ITD values is still larger and approaches a uniform distribution.

The effects of reverberation on the interaural level difference can also be substantial. An example here is based on recordings made from the two ears of KEMAR in a room in which all six surfaces may be covered with materials having different sound absorption properties. Figure 3 (from Kidd et al, 2005a) shows the variation of the measured ILD versus frequency for three room conditions. The room conditions are foam-covered ("FOAM"; near anechoic), typical untreated IAC booth surfaces ("BARE"; small amount of reverberation) or plexiglas-covered ("PLEX"; high reverberation). Note that there is a substantial reduction in the ILD as the reverberation increases, especially at high frequencies where the anechoic ILDs are larger. This reduction in ILD is also accompanied by a decrease in the peak of the cross-correlation function and in the direct to reverberant ratio.

The impulse responses for these three room conditions are shown in Figure 4. The recordings were made with a single microphone suspended at the approximate location of the center of the subject's head. It is obvious that the amount of reverberant energy increases steadily as the room surfaces become more reflective. The direct-to-reverberant ratio (D/R) is also given for each recording. Note that the most reflective case had a negative D/R ratio in decibels indicating that there
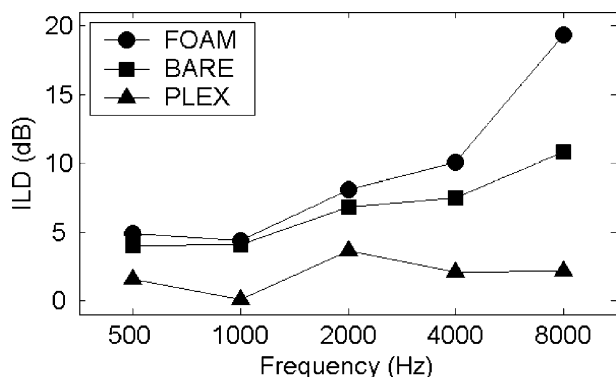
was more reflective energy than direct energy at the location of the head.

These observations illustrate that interaural differences are very complicated in reverberant environments. They vary substantially over time and depend in detail on the specific stimulus and the specific physical environment. It would be expected, therefore, that tests of functional hearing abilities in anechoic and reverberant environments may give different results. Furthermore, the effects of various amounts of reverberation could be different for listeners with normal hearing and listeners with hearing loss. In general, although most of daily living takes place in spaces that are reverberant and contain multiple sources of sound, there are few systematic studies of binaural abilities in realistic complex environments.

Complicated signals and informational masking

In addition to increasing the number of maskers or the amount of reverberation, the environment may increase in complexity as a result of the signals, and the relationships among the signals, becoming more complicated. Although under such circumstances the listener may be able to detect the presence of the target because there is only minor overlap of the target and maskers in frequency and time (i.e., there is negligible "energetic masking"), the listener may not be able to adequately identify or understand the target (e.g., a speech signal) because of an inability to properly segregate and attend to the target in the complex stimulus (i.e., there is substantial "informational masking").

It has been well-established that the amount of informational masking generally tends to be reduced when the target and masker are perceived to be spatially separated (e.g., Kidd et al, 1998; Freyman et al, 1999; Arbogast et al, 2002). In other words, spatial separation usually aids the listener in the tasks of segregation and of directing attention towards the target. Some results illustrating the reduction of informational masking due to the spatial separation of a speech target and masking sources are presented in Figures 6–8. The general strategy of the studies described next is to control the spectral overlap of the target and maskers in order to vary the relative amounts of energetic and informational masking. This is accomplished by reducing the stimuli into a set of very narrow frequency bands. The targets (and speech maskers) are a modified version of cochlear-implant-simulation speech (Shannon et al, 1995; refer to Arbogast et al, 2002, for details). Speech signals processed in this manner are highly intelligible and have well-controlled frequency content. These stimuli allow variation in the amount of energetic masking by presenting mutually exclusive (low energetic masking) or completely overlapping (high energetic masking) frequency bands for the target and the masker.

The first study is by Arbogast et al (2005), who compared spatial release from energetic and informational masking in listeners with normal hearing (NH) and sensorineural hearing loss (HI). There were three types of maskers – one comprised of speech and two comprised of noise – that were paired with the target sentences. The speech masker was processed into very narrow bands just like the target sentences, but consisted of frequency bands that were mutually exclusive with those of the target. This masker is referred to as "different-band sentence" or DBS. The different-band noise masker (DBN), was Gaussian

Figure 3. Interaural Level Differences (ILDs) as a function of frequency for three levels of reverberation in a sound-treated room. In order of increasing reverberation the rooms are labelled: FOAM, BARE and PLEX. From Kidd et al, 2005a.
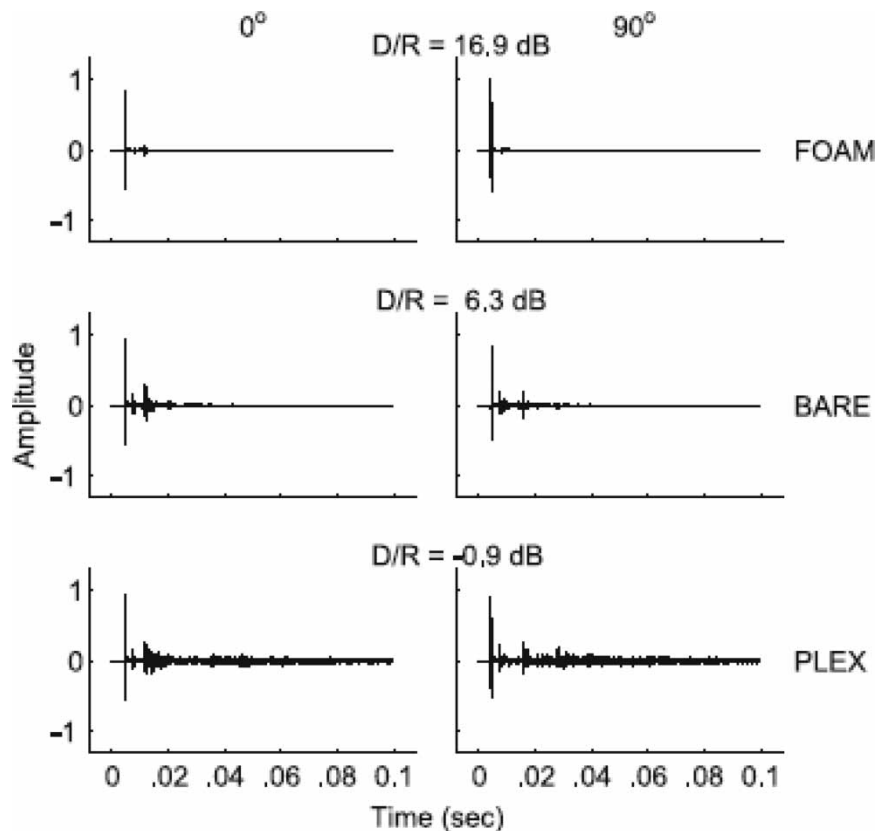
Figure 4. Impulse responses (IRs) for sound-treated room in three different reverberation conditions. The IRs were measured for the two speaker locations used in the spatial separation experiments (see text). From Kidd et al, 2005a.

noise processed into sets of equally narrow frequency bands that also were mutually exclusive with the target sentence bands. This masker was intended to provide an estimate of the small amount of energetic masking present for the DBS masker. The other noise masker – same-band noise or SBN – is comprised of a set of narrow bands of noise that directly superimpose on the target bands, thus maximizing energetic masking.

Examples of the magnitude spectra for a target paired with samples of each of these maskers is shown in Figure 5.

The sentences used were from the Coordinate Response Measure (CRM) corpus (Bolia et al, 2000) which have the format, "Ready [callsign], go to [color] [number] now." The task is to identify the color and number associated with a specified target callsign. The DBS masker always had a different callsign, color, and number than the target.

The measured values are speech reception thresholds (SRTs) computed as target-to-masker ratios corresponding to approximately 51% correct on the psychometric functions. The stimuli were played over loudspeakers in a sound-treated room. The location of the target loudspeaker was always straight ahead of the listener, and the location of the masker was either the same as the target (0 degrees) or separated by 90 degrees. The surfaces of the sound-treated room for this study were typical of a standard IAC booth with carpeted floor (condition BARE described above).

The results are shown in Figure 6, where the SRTs are plotted for all of the conditions described above.

The three masker types are displayed along the abscissa. For each masker, the two spatial separations are shown, and, for each separation, the average results are plotted for both listener groups. First, the SRTs are lower for the NH group for all different-band conditions. The difference between DBS- and DBN-masked SRTs, which is taken as an estimate of the additional informational masking produced by the DBS masker, was greater in the NH group. This may be because the wider filters in the HI group resulted in more direct interaction between target and masker bands causing the energetic component to dominate the listening situation. Second, there are no significant differences between groups for the SBN masker. This result suggests that there are no differences in on-band energetic masking between groups. Third, the SRTs for the DBN masker are lower than for either of the other maskers for both groups and spatial separations. That result was expected given that the DBN masker produces relatively little masking of any kind. The difference between the SRTs for the DBN and SBN maskers is related to the width of the auditory filters and clearly distinguishes the two groups. For 0-degree separation, the difference was about 25 dB for the NH group and only about 14 dB for the HI group.

The spatial release from masking apparent in Figure 6 (lower SRTs for the 90-degree than for the 0-degree separation) is plotted in Figure 7. For both listener groups, the largest spatial release from masking is observed for the highly informational DBS masker. This release of nearly 10 dB for the HI group and
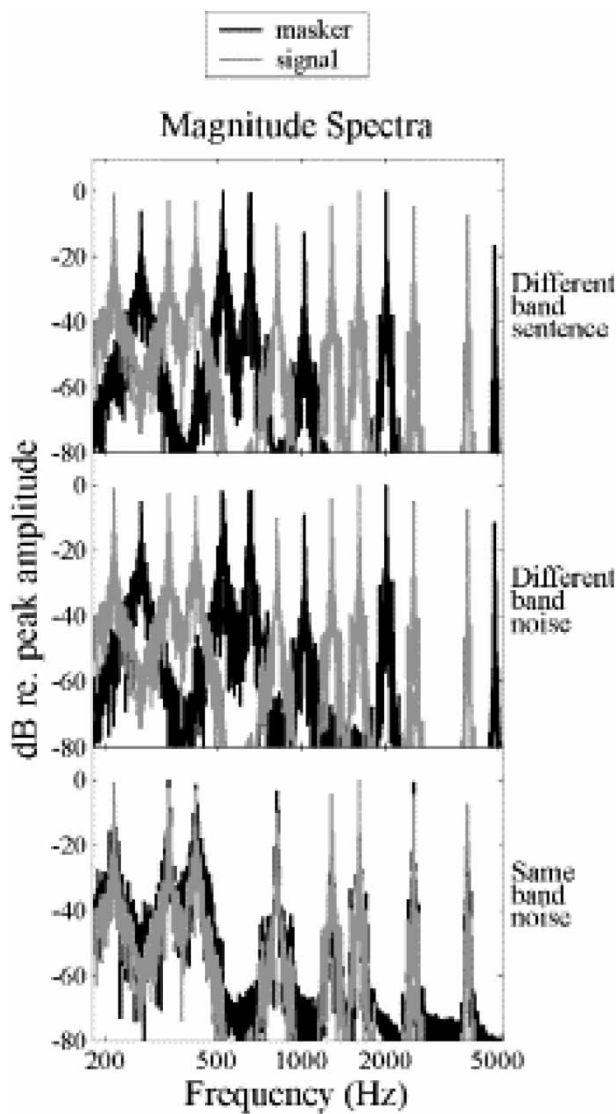
Figure 5. Magnitude spectra for a target (grey) and three types of maskers (black): different-band speech, different-band noise and same-band noise. Figure from Arbogast et al, 2005.

almost 15 dB for the NH group is significant in that: 1) it demonstrates that a large release from masking due to spatial separation of sources may be obtained because of perceptual factors, rather than traditional binaural analysis, and 2) it confirms that spatial-perceptual cues may provide a large benefit to listeners with sensorineural hearing loss in complex, multi-source listening environments.

EFFECTS OF REVERBERATION ON SPATIAL RELEASE FROM MASKING

A similar approach to that described above was used to explore the effects of reverberation on spatial release from masking by Kidd et al (2005). Only NH listeners were tested, but the conditions and stimuli were nearly the same as in Arbogast et al (2005). The primary addition in this new study was the variation in the reflectivity of the surfaces of the room. As described earlier in this chapter, panels of foam or of Plexiglas were
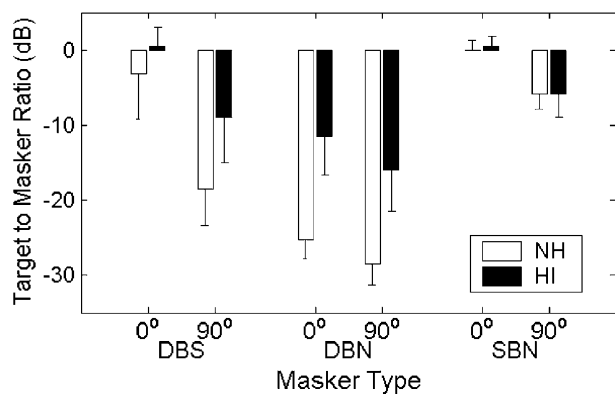

Figure 6. Speech reception thresholds for different conditions plotted as target-to-masker ratios at 51% correct performance. The abscissa displays the three different masker types and the two spatial separation conditions. The key identifies normal hearing (NH) and hearing-impaired (HI) group mean data. Figure is adapted from Arbogast et al, 2005.

attached to all surfaces of a large single-walled IAC sound booth, and SRTs were measured for both spatial separation conditions (0 and 90 degrees). One methodological difference was that SRTs were measured using an adaptive tracking procedure.

The main focus of this study is the interaction between the masker type, which is believed to be correlated with whether the dominant masking is informational or energetic, and the degree of reverberation. As the degree of reverberation increases, the later-arriving components of the sounds are increasingly diffuse. That is, the reflections come from increasingly diverse directions, superimposing upon each other, so that the sounds at the two ears are more and more decorrelated. The consequence of this is that there is correspondingly less of an advantage available due to binaural processing.

The results of this study of the effects of reverberation are presented in Figure 8. Several points can be noted. First, the amount of spatial release from masking for the DBN condition is relatively small in all conditions but did decrease with increased reverberation. The small release is consistent with a
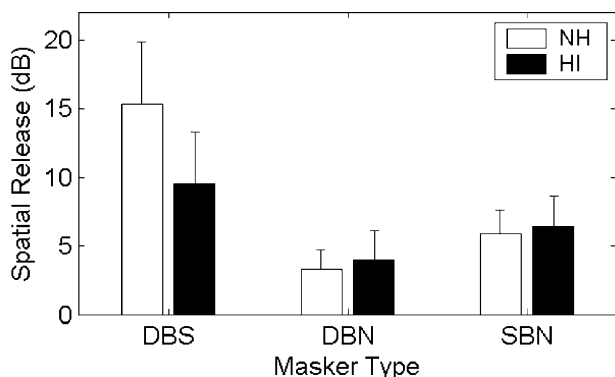

Figure 7. Spatial release from masking for three masking conditions for both normal-hearing (NH) and hearing-impaired (HI) listeners. Listening conditions are described in the text. Figure is adapted from Arbogast et al, 2005.
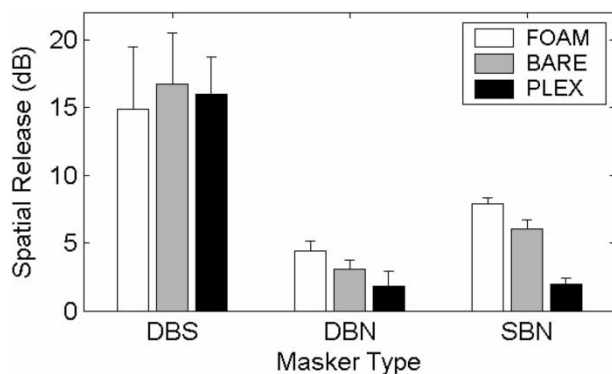
Figure 8. Group mean spatial release from masking, and standard errors of the means, for DBS, DBN, and SBN maskers tested in FOAM, BARE, and PLEX room conditions (see text). Figure is adapted from Kidd et al, 2005.

small amount of masking to begin with as discussed above. Second, the amount of masking in the SBN condition shows about eight decibels of spatial release in near-anechoic conditions, and this advantage decreases to only about two decibels as the reverberation increases (c.f., Plomp, 1976; Zurek, 1993). Because, as we have argued, the interaural correlation of the signals at the two ears is decreasing as the reverberation increases, the usual mechanisms of binaural processing become less effective. Thus, only the better ear advantage is expected to produce significant spatial release from masking in a reverberant room. Moreover, if the room is highly reverberant, the energy reaches the head from all directions (and this spatial distribution does not change significantly with the direction of the source relative to the listener since the direct sound accounts for only a small portion of the total energy). In other words, because of the omnidirectional nature of reverberant energy, the better ear advantage diminishes as the reverberation increases. Finally, note that the DBS condition maintains a substantial spatial advantage even with high reverberation. The spatial release in this case is largely unaffected by the level of reverberation, at least over the range of values used in this experiment. The explanation suggested by Kidd et al (2005), supported by the results of experiments by Freyman et al (1999, 2001), is that the precedence effect allows the two sources to be localized independently and that the separate perceived locations allow a relatively easy resolution of any confusions about which source produced the target test words.

Summary and concluding remarks

Much of binaural processing can be understood by a consideration of the processing of interaural differences in time and level. These interaural differences are available for each frequency band; they are based on fine structure at low frequencies and envelopes at high frequencies; and they are best characterized as sequences of temporal values. Listeners are sensitive to the width of the distribution of these values as well as to the mean value, so that interaural differences provide a basis for binaural detection and interaural correlation sensitivity as well as for lateralization and interaural discrimination.

In complex acoustical environments (with reverberation and multiple sources), these interaural differences are complex patterns and are interpreted through a complex process that includes both top-down and bottom-up factors.

Since many of the environments that listeners normally encounter are acoustically complex and since there are clear advantages to binaural listening, it may be useful to develop evaluation procedures that could be used for diagnosis and for evaluation of auditory prostheses (e.g., hearing aids and cochlear implants).

At a more general level, knowledge about binaural processing and binaural perception has a wide variety of applications in the clinical domain. Generally speaking, such knowledge has implications for audiological testing, for behavioral training of listeners with impaired hearing, and for the development of improved hearing aids. To this time, measurements of binaural capabilities do not comprise a significant part of the testing at most clinics although there are some obvious advantages of developing such tests. Similarly, the training area is not pursued as yet. Of greater importance and current usefulness, knowledge about binaural perceptual capabilities and about how the binaural system works provides important background for the development of improved hearing aids. For listeners with sufficient residual hearing in the two ears to make binaural aids worthwhile, consideration must be given to such issues as automatic gain control and to how such control can be usefully integrated with binaural function. For listeners with one ear, knowledge about the normal binaural system can inform design of monaural aids that make use of an array of microphones to achieve spatial resolution and special encoding schemes to encode source direction monaurally. In general, the extent to which such artificial multi-sensor systems can substitute for complete loss of hearing in one ear (or even perhaps surpass the normal binaural system in some important ways) remains to be seen.

References

Arbogast, T.L., Mason, C.R. & Kidd, G. Jr. 2002. The effect of spatial separation on informational and energetic masking of speech. *J Acoust Soc Am*, 112, 2086–2098.

Arbogast, T.L., Mason, C.R. & Kidd, G. Jr. 2005. The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 117, 2169–2180.

Bernstein, L.R., Trahiotis, C. & Hyde, E.L. 1998. Interindividual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise. *J Acoust Soc Am*, 103, 2069–2078.

Bolia, R.S., Nelson, W.T., Ericson, M.A. & Simpson, B.D. 2000. A speech corpus for multitalker communications research. *J Acoust Soc Am*, 107, 1065–1066.

Boudreau, J.C. & Tsuchitani, C. 1970. Cat superior olive S-segment cell discharge to tonal stimulation. In W.D. Neff (ed.), *Contributions to Sensory Physiology*. Vol. 4 New York: Academic Press, pp. 143–213.

Durlach, N.I. & Colburn, H.S. 1978. In E. Carterette & M. Friedman (eds.) *Binaural Phenomena, chapter in Handbook of Perception: Hearing*. Vol. 4 Academic Press: New York.

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

Freyman, R.L., Helfer, K.S., McCall, D.D. & Clifton, R.K. 1999. The role of perceived spatial separation in the unmasking of speech. *J Acoust Soc Am*, 106, 3578–3588.

Freyman, R.L., Balakrishnan, U. & Helfer, K.S. 2001. Spatial release from informational masking in speech recognition. *J Acoust Soc Am*, 109, 2112–2122.

Freyman, R.L., Balakrishnan, U. & Helfer, K.S. 2004. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am*, 115, 2246–2256.

Gabriel, K.J. & Colburn, H.S. 1981. Interaural correlation discrimination: I Bandwidth and level dependence. *J Acoust Soc Am*, 69, 1394–1401.

Goldberg, J.M. & Brown, P.B. 1969. Response of Binaural Neurons of Dog Superior Olivary Complex to Dichotic Tonal Stimuli: Some Physiological Mechanisms of Sound Localization. *J Neurophys*, 32, 613–636.

Goupell, M.J., & Hartmann, W.M. (2006). Interaural fluctuations and the detection of interaural incoherence. I. Bandwidth effects, *J Acoust Soc Am*, in press.

Grothe, B. 2003. New roles for synaptic inhibition in sound localization. *Nat Rev Neurosci*, 4, 540–550.

Guinan, J., Guinan, S.S. & Norris, B.E. 1972. Single auditory units in the superior olivary complex I: Responses to sounds and classifications based on physiological properties. *Intern J Neuroscience*, 4, 101–120.

Harper, N. & McAlpine, D. 2004. Optimal neural population coding of an auditory spatial cue. *Nature*, 430, 682–686.

Hausler, R., Colburn, H.S. & Marr, E. 1983. Sound localization in subjects with impaired hearing. *Acta Otolaryngol (Stockh)*, Suppl 400, 1–62.

Hawley, M.L., Litovsky, R.Y. & Colburn, H.S. 1999. Speech intelligibility and localization in a multi-source environment. *J Acoust Soc Am*, 105, 3436–3448.

Henning, G.B. 1974. Detectability of interaural delay in high-frequency waveforms. *J Acoust Soc Am*, 55, 84–90.

Jeffress, L.A. 1948. A place theory of sound localization. *J Comp Physiol Psychol*, 41, 35–39.

Joris, P.X., Carney, L.H., Smith, P.H. & Yin, T.C. 1994. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *J Neurophysiol*, 71, 1022–1036.

Joris, P.X., Smith, P.H. & Yin, T.C. 1998. Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron*, 21, 1235–1238.

Kidd, G., Jr., Mason, C.R., Rohtla, T.L. & Deliwala, P.S. 1998. Release from masking due to spatial separation of sources in the identification of non-speech auditory patterns. *J Acoust Soc Am*, 104, 422–431.

Kidd, G. Jr., Mason, C.R., Brughera, A. & Hartmann, W.M. 2005. The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica with Acustica*, 91, 526–536.

Koehnke, J., Colburn, H.S. & Durlach, N.I. 1986. Performance in several binaural-interaction experiments. *J Acoust Soc Am*, 79, 1558–1562.

Langendijk, E.H., Kistler, D. & Wightman, F.L. 2001. Sound localization in the presence of one or two distracters. *J Acoust Soc Am*, 109, 2123–2134.

Moore, J.K., Simmons, D.D. & Guan, Y-L. 1999. The Human Olivocochlear System: Organization and Development. *Audiology & Neuro-Otology*, 4, 311–325.

Nuetzel, J.M. & Hafter, E.R. 1981. Discrimination of interaural delays in complex waveforms: Spectral effects. *J Acoust Soc Am*, 69, 1112–1118.

Peissig, J. & Kollmeier, B. 1997. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *J Acoust Soc Am*, 101, 1660–70.

Plomp, R. 1976. Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acustica*, 34, 200–211.

Pollack, I. & Trittipoe, W.J. 1959. Binaural listening and interaural noise cross correlation. *J Acoust Soc Am*, 31, 1250–1252.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J. & Ekelid, M. 1995. Speech recognition with primarily temporal cues. *Science*, 270, 303–304.

Shinn-Cunningham, B.G., Santarelli, S. & Kopco, N. 2000. Tori of confusion: Binaural localization cues for sources within reach of a listener. *J Acoust Soc Am*, 107, 1627–1627-1636.

Shinn-Cunningham, BG and Kawakyu, K. (2003). Neural representation of source direction in reverberant space, *In: Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Pfaltz, NY, 19–22 October 2003.

Shinn-Cunningham, B.G. (2005). Influences of spatial cues on grouping and understanding sound, *In: Proceedings of Forum Acusticum*, 29 August-2 September.

Yin, T.C.T. & Chan, J.C.K. 1990. Interaural Time Sensitivity in the Medial Superior Olive of the Cat. *J Neurophys*, 64, 465–488.

Zurek, P.M. 1991. Probability distributions of interaural phase and level differences in binaural detection stimuli. *J Acoust Soc Am*, 90, 1927–1932.

Zurek, P.M. 1993. Binaural advantages and directional effects in speech intelligibility. In G.A. Studebaker & I. Hochberg (eds.) *Acoustical Factors Affecting Hearing Aid Performance*. Allyn and Bacon: Boston, pp. 255–276.