Perception & Psychophysics
2004, 66 (8), 1379–1387

Some characteristics of auditory spatial attention revealed using rhythmic masking release

ANDREW J. SACH and PETER J. BAILEY
University of York, York, England

The tuning of auditory spatial attention with respect to interaural level and time difference cues (ILDs and ITDs) was explored using a rhythmic masking release (RMR) procedure. Listeners heard tone sequences defining one of two simple target rhythms, interleaved with arhythmic masking tones, presented over headphones. There were two conditions, which differed only in the ILD of the tones defining the target rhythm: For one condition, ILD was 0 dB and the perceived lateral position was central, and for the other, ILD was 4 dB and the perceived lateral position was to the right; target tone ITD was always zero. For the masking tones, ILD was fixed at 0 dB and ITDs were varied, giving rise to a range of lateral positions determined by ITD. The listeners' task was to attend to and identify the target rhythm. The data showed that target rhythm identification accuracy was low, indicating that maskers were effective, when target and masker shared spatial position, but not when they shared only ITD. A clear implication is that at least within the constraints of the RMR paradigm, overall spatial position, and not ITD, is the substrate for auditory spatial attention.

Over a century ago, William James (1890) suggested that focusing attention might be accomplished by muscular changes in the appropriate sense organ. Such early selection at the receptor would lift from the central nervous system much of the burden of extracting target information from competing stimuli. This process is a familiar aspect of overt orienting with vision, in which eye movements control which objects in a visual scene fall on the fovea. In certain animals with prominent and highly directional pinnae, head and pinna movements may perform a similar function for spatial focusing of audition. In humans, this effect is limited, and since the cochlea lacks an equivalent to the fovea, it is less obvious how attentional selection of auditory spatial location might be implemented early in the peripheral auditory system.

Information about the azimuthal position of a sound source is encoded by the auditory system in two primary ways: (1) for any source displaced from the sagittal plane, the sound reaches the closer ear earlier than it does the farther one, giving rise to an interaural time difference (ITD) cue. (2) Particularly for high frequencies, the acoustic shadow cast by the head results in the sound being more intense at the closer ear, yielding an interaural level difference (ILD).[1] This article is concerned with the tuning of auditory spatial attention with respect to these cues and with whether directing auditory attention to some spatial position defined only by an ILD is equivalent to directing attention to the same position defined by an ITD. This question is closely related to whether auditory spatial attention operates relatively early or late in the auditory pathway, as discussed below.

A potential substrate for early attentional selection by spatial location was described in a visionary paper by Jeffress (1948). He proposed a neural mechanism in the auditory pathway for the detection of ITDs, whereby an array of central units respond to coincidences in the signals arriving from the two ears via fibers of different lengths, which function as delay lines, so that each "coincidence detector" is tuned to a particular ITD. Jeffress's proposal, which has since found physiological support (Goldberg & Brown, 1969; Yin & Chan, 1990), raises the possibility that listeners could selectively monitor those coincidence detectors associated with a particular target azimuth and thus provides a potential mechanism for spatial attention involving low-level selection by ITD. Elaborations of such a scheme have been implemented in computational approaches to auditory scene analysis (e.g., Bodden, 1996).

Figure 1a is a schematic representation of the auditory system in which ITD and ILD cues are processed separately before the two are combined to yield an integrated spatial percept. This later integration stage may also incorporate cues from other modalities, such as vision (see, e.g., Driver, 1996; Summerfield & McGrath, 1984). It was once thought that information from ILDs and ITDs was effectively combined as early as the auditory nerve (Deatherage & Hirsh, 1959), but converging evidence for parallel and independent decoding of ITD and ILD cues has now emerged from a number of sources, including human neuropsychology (e.g., Griffiths, Elliott, Coulthard, Cart-

---

Financial support was provided by the U.K. Biotechnology and Biological Sciences Research Council, U.K. Medical Research Council, and (through a travel bursary to A.J.S.) the British Psychological Society. The contributions of Al Bregman, Martine Turgeon, Pierre Ahad, Nick Hill, David Wellsted, Philip Quinlan, and Chris Darwin are gratefully acknowledged. Correspondence concerning this article should be addressed to P. Bailey at the Department of Psychology, University of York, York YO10 5DD, England (e-mail: pjb1@york.ac.uk).

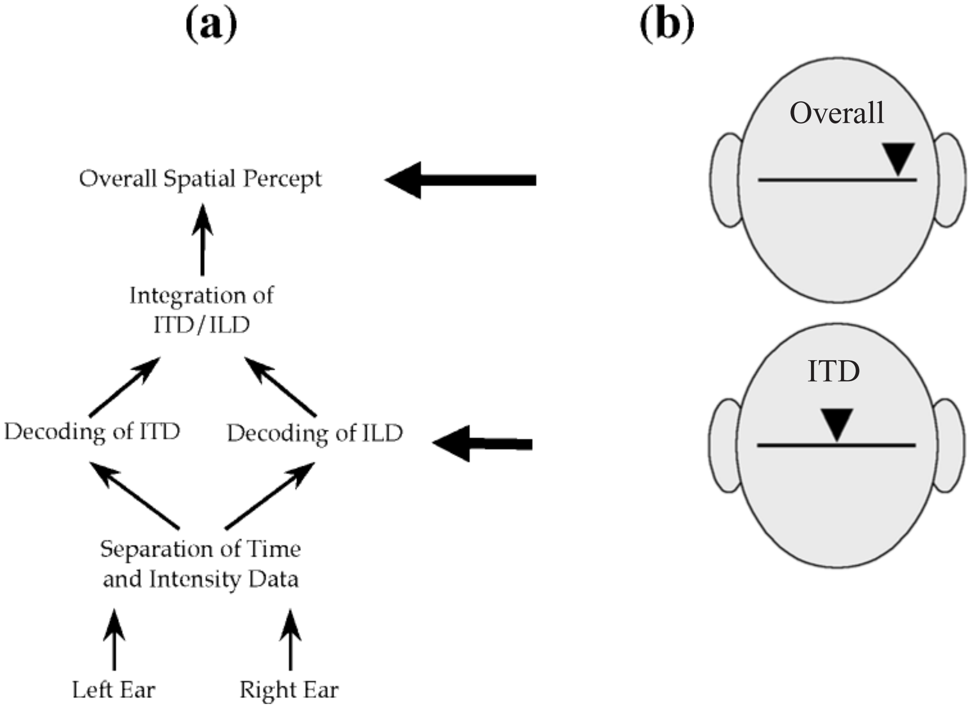

Figure 1. (a) Parallel decoding of ITD and ILD cues in the auditory system. (b) Diagram indicating, for a target with ITD = 0, ILD = 4 dB, the two lateral positions at which a masker is expected to have the most deleterious effect on rhythm identification, depending on where in the pathway attentional selection takes place. The positions are shown on a pair of schematic coronal sections through the head. If attention occurs early, in the ITD pathway, then the most effective maskers should be those for which ITD matches the target ITD—that is, those in the center (ITD = 0). If attention occurs late, after integration of ITDs and ILDs, then the most effective maskers should be those for which ITD matches the overall spatial position of the target—that is, those to the right.

lidge, & Green, 1998), human psychophysics (e.g., Gilliom & Sorkin, 1972; Hafter & Jeffress, 1968), human electrophysiology (e.g., Schröger, Tervaniemi, Winkler, Wolff, & Näätänen, 1997), and nonhuman electrophysiology (e.g., Olsen, Knudsen, & Esterly, 1990).

If this scheme is correct, a demonstration that systematic differences exist between the effects of attention to ITDs and to ILDs would constitute strong evidence that attention operates relatively early, at a level of processing before the two binaural cues are combined. If, however, ITDs and ILDs produce identical effects, then it would be economical to argue that they are equivalent, in line with the hypothesis that attention operates relatively late, after ITDs and ILDs have been integrated to yield a single perceived locale.

Darwin and Hukin (1999) suggested that the "early selection" account is not supported by evidence from experiments on auditory perceptual grouping. They argued that a model of spatial attention in which listeners select a column of coincidence detectors corresponding to a common ITD across frequency predicts that a common ITD should be an effective basis for grouping simultaneous sounds at different frequencies. However, common ITD has typically been shown to be a weak basis for simultaneous auditory grouping (e.g., Culling & Summerfield, 1995; Darwin & Hukin, 1997), although it may be usable by highly trained listeners (Drennan, Gatehouse, & Lever, 2003). Darwin and Hukin (1999) therefore contended that the limited role of ITD in simultaneous perceptual grouping indicates an attentional process that operates later in the auditory pathway, at a stage when auditory objects have been formed and localized. This argument is consistent with Bregman's (1990) suggestion that "streams are formed by some other process, not the one we call attention, but that attention can select one of these already formed streams for further processing." (p. 192).[2] Darwin and Hukin's view that attention is directed toward "subjective locations" rather than toward ITDs alone has implications for the relative contribution of different spatial cues to an attentional mechanism: Once the direction of an auditory object is fixed, prior to attention, the individual interaural cues that contribute to estimates of that direction become indistinguishable, so that attention to ITD and to ILD amount to the same thing.

In this study, we describe a new paradigm for exploring auditory spatial attention, or at least a new application of an existing one. Using the phenomenon of rhythmic masking release described below, we measured the ITD-tuning characteristic of spatial attention. The paradigm involved the use of both an attended target stream and unwanted distractors (maskers); independent manipulation of the ILD and ITD of the target and maskers, respectively, facilitated a test of the "early" and "late" hypotheses outlined above.

Probe–signal studies of auditory spatial attention have commonly used only two spatial positions, the minimum required for comparison between attended and unattended (probe) locations (e.g., Mondor & Amirault, 1998; Mondor & Breau, 1999; Sach, Hill, & Bailey, 2000; Spence & Driver, 1994). Studies of auditory attention in the frequency domain have conversely tended to use a number of probe frequencies, allowing the tuning characteristic (or "listening band") of attention to be mapped more explicitly. There is much to be gained by the latter, more fine-grained approach. First, multiprobe data are more compelling as a demonstration of the reality of attentional processes, in that they are not so easily rationalized as a task-specific epiphenomenon.[3] Second, the shape of the tuning curve traced out by multiple probes (and in particular its width) may give clues as to the neural substrate of the attentional filter. On the basis of such data in the frequency domain, Scharf (1998) argued for identification of the listening band with the auditory filter itself, showing that attention must select between those frequency channels already established at the cochlea. There might be more to learn about auditory spatial attention if such quantitative information about filtering were available. Rather than testing a variety of probe locations for a single expected/cued location, the present study held the target location constant while tracing out the attentional focus using maskers with a range of ITDs.

The phenomenon of rhythmic masking release (RMR) was first presented as an auditory demonstration (refer to track 22 of Bregman & Ahad, 1996). It involved presentation of a sequence of pure tones with an isochronous rhythm that was camouflaged by irregularly spaced masking tones of the same frequency interleaved with the targets. The target rhythm was not distinguishable until flanker tones of a different frequency (separated by $>$ 1 critical band) were added synchronously with the maskers. The maskers and flankers fused perceptually to form an independent stream with a timbre different from the targets. This integration of simultaneous components took precedence over the sequential grouping of targets and maskers, and accordingly the target rhythm was released from masking. Thus, adding flanker energy synchronous with the maskers had the effect of *decreasing* the amount of masking. Bregman and Ahad noted that in this respect, RMR is related to the phenomenon of comodulation masking release (Moore, 1990).

RMR was developed into a useful paradigm for the study of the cues important for simultaneous auditory grouping by Turgeon (Turgeon, 1999; Turgeon, Bregman, & Ahad, 2002). Rather than using a single, isochronous rhythm, Turgeon had listeners distinguish between two possible rhythms in a two-alternative forced choice procedure. She then manipulated the relationship between the maskers and the flankers, experimenting with variables such as onset asynchrony, harmonicity, and spatial position, reasoning that the conditions in which listeners performed best at the rhythm discrimination were those for which the maskers and flankers showed the strongest tendency to group together.

In the present experiment, the flanker tones were dispensed with. Whereas in the original paradigm, the targets and maskers were always identical, the innovation here was to separate the two in terms of cues relating to lateral spatial position, so that the ease with which listeners "heard out" the target rhythm could be measured as the cues defining the lateral positions of targets and maskers were systematically manipulated. The task was attentional in that listeners must attempt to focus on (attend to) the target position while ignoring all sounds from other locations. This task modeled a real-world cocktail party in which the listener wishes to monitor a single conversation and to suppress sounds from other locations. However, at a cocktail party, several cues would aid segregation of and attention to the target voice (such as pitch, idiosyncratic formant patterns associated with different vocal tracts, prosody, linguistic context, etc.), but in this experiment there was only one: The success of attentional filtering depended exclusively on the precision with which listeners could focus on one point in lateral space to the exclusion of others. This approach to studying spatial attention as a *filtering* process, as distinct from one that focuses on attentional *control* by contrasting patterns of performance in attended and unattended conditions, follows the tradition begun with the early studies of dichotic listening (e.g., Broadbent, 1958; Cherry, 1953), in which a listener had to monitor a target channel in the presence of distractor sounds in a competing channel. Attention in this scheme is understood as the measure of success with which one can successfully monitor the target channel and overcome the deleterious interference of the distractors. Although their study was not specifically concerned with attention, Kidd, Mason, Rohtla, and Deliwala (1998) have used a similar approach in their experiments on release from informational masking resulting from spatial separation of sound sources.

The RMR paradigm was used here to explore whether attention to ITDs and to ILDs are equivalent. Masker tones were given a range of ITDs but always had an ILD of 0 dB. Thus, the maskers traced out the tuning characteristic of spatial attention by means of ITD exclusively. The focus of the experiment was to discover whether this characteristic is any different from the tuning characteristic with respect to *overall spatial position.* That is, does the tuning function relate to an attentional substrate specific to ITDs or to one that deals in terms of the overall spatial position, regardless of whether that is defined by ITDs or ILDs? To answer this question, two experimental conditions were defined in which the target, for which the ITD was always zero, had an ILD of either 0 or 4 dB. When the target ITD = ILD = 0, the tuning function would be expected to center on ITD = 0; maskers should be most deleterious to rhythm identification when they are closest to the targets, but proximity in terms of ITD and of overall spatial position amount to the same thing in this condition. The essence of the experiment is the comparison between this ITD = ILD = 0 condition and the other, in which the target ITD is 0 but ILD is 4 dB. When target ILD is 4 dB, the overall spatial position of the target is shifted to one side (although its ITD remains at zero).[4] If attention operates early, and within the ITD pathway, then the tuning curve should be centered as before at ITD = 0. However, if attention operates late, on the overall spatial position determined by combining ITD and ILD cues, then the tuning curve should shift along with the perceived target position. These possibilities are illustrated in Figure 1b.

METHOD

Participants

Seven female and 3 male listeners aged between 18 and 22 years were either paid or received course credit for their participation in the experiment. The listeners were divided equally into two groups, each of which experienced the two experimental conditions in a different order. All listeners had normal pure tone thresholds in the range tested (250–8000 Hz).

Stimuli and Apparatus

Target and masker tones were 500-Hz sinusoids of 50-msec duration, including 5-msec cosine-squared onset and offset ramps. For an ILD of zero, the monaural level was measured to be 80 dB SPL. Masker ITDs were realized by manipulating the interaural phase difference. Amplitude envelopes were synchronous at the two ears. Following Hill and Darwin (1996), the ILD of 4 dB was applied asymmetrically to avoid loudness differences between tones with different ILDs.

Stimuli were synthesized in advance at a sampling rate of 22.05 kHz. The RMR sequences were constructed in real time using the MCF software by DigiVox[5] running on an IBM-compatible PC. The resulting waveforms were converted to voltages using 24-bit DACs (LynxOne). Stimuli were presented over Sennheiser HD580 headphones. Each listener was tested individually in a sound-attenuating enclosure.

Structure of RMR Sequence

The listeners' task was to discriminate between two target rhythms, as indicated in Figure 2a. Both of these rhythmic sequences were constructed using equal numbers of short (300-msec) and long (600-msec) intertarget intervals (ITIs, defined as the interval between the offset of one target and the onset of the next). The sequences differed only in the pattern of these two ITIs, so that the mere awareness of a particular temporal interval present within the sound was insufficient to establish the rhythm.

Masker tones, identical to the targets except for the value of interaural parameters responsible for lateralizing the sounds within the head, were interspersed with the targets to confound discrimination of the target rhythm. Two such maskers were inserted into each short ITI, and four into each long ITI. A strategy was devised for determining the timing of the maskers within an ITI that was intended to allow for a high degree of variability in masker position, while preventing the "bunching up" of maskers that would be likely if their timing was completely random. Allowing for a minimum of 10 msec to separate any two successive sounds, the remaining ITI was divided up equally into two or four temporal windows, for short and long ITIs, respectively. Thus, each window within a short ITI was 135 msec long, and each within a long ITI was 137.5 msec long. Each masker was constrained to fall within one of these windows, but its temporal position was otherwise random. These constraints are summarized in Figure 2b. The sequence began with one of the maskers at random, to prevent listeners from identifying the target rhythm relative to the first sound heard. A sequence continued cyclically until terminated by a response, or for a maximum of 8 sec.

Procedure

There were two conditions, in which the target ILD was either 0 or 4 dB. Targets were lateralized to the right when the ILD was 4dB. The ITD of the targets was always zero. The conditions were counterbalanced across listeners, so that half of the listeners experienced the target ILD = 0 dB condition followed by the ILD = 4 dB condition, and vice versa.

The ILD of the maskers was always zero. Conversely, the masker ITD was an independent variable having 10 levels (±600, ±350, ±200, ±100, ±50 μsec). A positive ITD was lateralized to the right. For each level of masker ITD, there were 2 levels of the target rhythm, A and B. The resulting 20 trial types were randomized to create a different order for each block of 20 trials, with one occurrence of each trial type every 20 trials.

There were 400 trials per ILD condition, with a short break (minimum 20 sec) separating every 100 trials. A longer break of several minutes separated the two conditions. Each listener was tested during a single session of approximately 90-min duration.

The two-alternative forced choice task was to identify which of the two target rhythms was presented on each trial. Listeners indicated a response by pressing a key, an action that also terminated presentation of the rhythmic sequence. Listeners were instructed to respond as quickly as possible while trying to minimize errors. There was no time limit on responding, although in the absence of a keypress the rhythmic sequence self-terminated after 8 sec.

A visual display showed a schematic representation of the two rhythms, similar to that given in Figure 2a; the display also indicated the number of trials completed. Right/wrong feedback immediately followed the response and lasted for 500 msec. A further 500 msec elapsed prior to the beginning of the next trial.

RESULTS

Plotting the proportion correct on the rhythm identification task as a function of the ITD of the masker tones yielded a spatial "tuning curve," analogous in some respects to the "listening band" functions used to characterize the results of probe–signal studies of auditory attention in the frequency domain (Scharf, 1998). As argued above, the listeners' task of focusing on the target at the expense of the maskers was essentially one of attentional filtering. Thus, the graphs may be thought of as *attentional* tuning curves.

Results for each of the listeners are plotted separately in Figures 3 and 4, the former displaying data for the listeners who completed the target ILD = 0 dB condition first. Results from the two conditions are displayed in adjacent panels. *Exposure Times* refer to the mean length of the RMR sequence that a listener chose to listen to before making a response. A summary of the results across listeners is given in Figure 5.

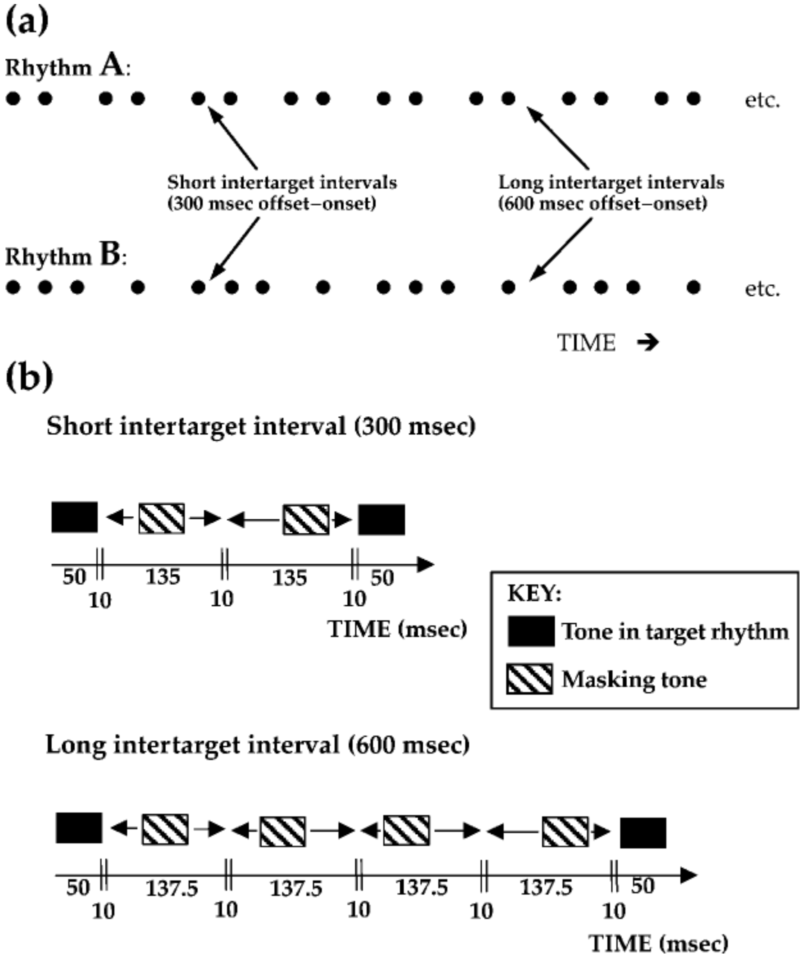

Figure 2. (a) Temporal structure of the two target rhythms. (b) Constraints on masker tone position within intertarget (offset–onset) intervals, to prevent masker tones from "bunching up"; filled boxes denote a tone in the target rhythm, cross-hatched boxes denote a masker tone free to adopt any position within its temporal window, indicated by the arrows.

DISCUSSION

Considering first data from the target ILD = 0 condition, all listeners showed a spatial tuning curve centered about a masker ITD of 0. The consistency across listeners is reflected in the small standard errors in Figure 5. As expected, maskers far away from the target ITD were easily ignored and had little effect on the rhythm discrimination. In contrast, maskers close to the target ITD were more distracting and considerably reduced listeners' ability to "hear out" the target rhythm. For the majority of listeners, the range of positions over which maskers had an appreciably deleterious effect was small, reflecting a sharp tuning. In an attempt to quantify the sharpness of auditory spatial tuning, we calculated something analogous to the "equivalent rectangular bandwidth" measure used to describe frequency tuning.[6] This calculation involved determining the width of a rectangle of unit height that bounds the same area as the averaged tuning curve (from Figure 5). The bandwidth of the spatial tuning curve estimated by this procedure was 337 μsec. The ITD detection threshold of 17 μsec at 500 Hz (Klumpp & Eady, 1956) is much smaller than this, in accord with what has been found for the frequency domain: Frequency discrimination thresholds are smaller than the auditory filter bandwidth and similarly smaller than typical estimates of the bandwidth of the attentional tuning curve for frequency, as revealed by probe–signal procedures. In terms of audible angle, 337 μsec span the segment -20° to $+20^\circ$, which is an impressively small range, but again compares poorly with the minimum audible angle (MAA) of approximately 1° at 500 Hz (Mills, 1958). The ability to attend to one particular ITD as against another, however, clearly requires more than the discrimination of the two, so perhaps a fairer comparison is the *concurrent* MAA, measured for two simultaneous sounds.[7] Depending on the spectral re-

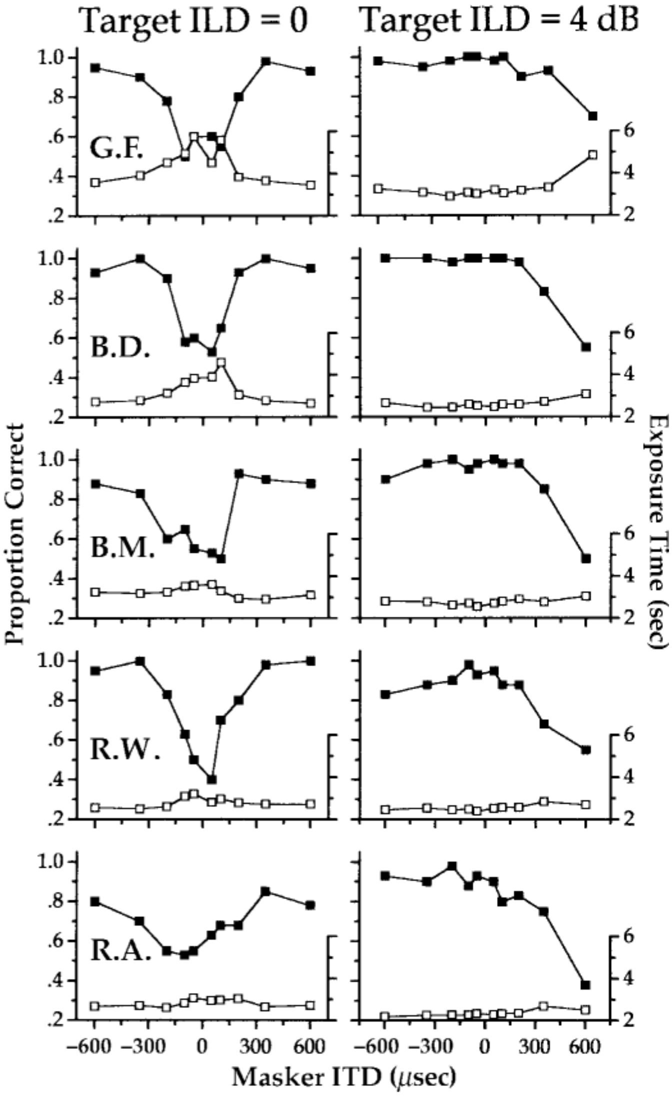

Figure 3. Results for Listeners 1–5. The left-hand panels show data from the target ILD = 0 condition, which for these listeners was experienced first. The right-hand panels display data from the target ILD = 4 dB condition. The ITD of the masker tones is plotted on the abscissa. Proportion correct (filled symbols) and exposure times (open symbols) are plotted on the ordinate; note the different scales.

solvability of the two sounds, this measure ranges between 5° and 46° (Perrott, 1984).

A few other studies have attempted to measure the shape or the width of the attended spatial focus. The closest to the present study in terms of methodology was that of Bronkhorst and Plomp (1988). In a simulated cocktail party experiment, they calculated the speech reception thresholds for a range of noise azimuths when ITD was the only spatial cue. The resultant spatial tuning characteristic was rather wider than that found in the present study. Similarly, Mondor and Zatorre (1995) found a comparatively shallow attentional gradient using a multiprobe–signal task with various angular separations between the cue and probes and using response latency as the dependent variable. Conversely, Teder and Näätänen (1994) reported a much narrower focus; a feature of their event-related potential data associated with sounds at an attended locus was found to fall off appreciably for sounds only 3° away (see also Röder, Teder-Sälejärvi, Sterr, Rösler, Hillyard, & Neville, 1999). Despite the variety of these results, they at least make clear that spatial attention is capable of finer resolution than the crude categories of left and right.

Few listeners showed a systematic variation in exposure time with masker ITD, but for those who did, the pattern was always the reverse of that seen in the proportion cor-

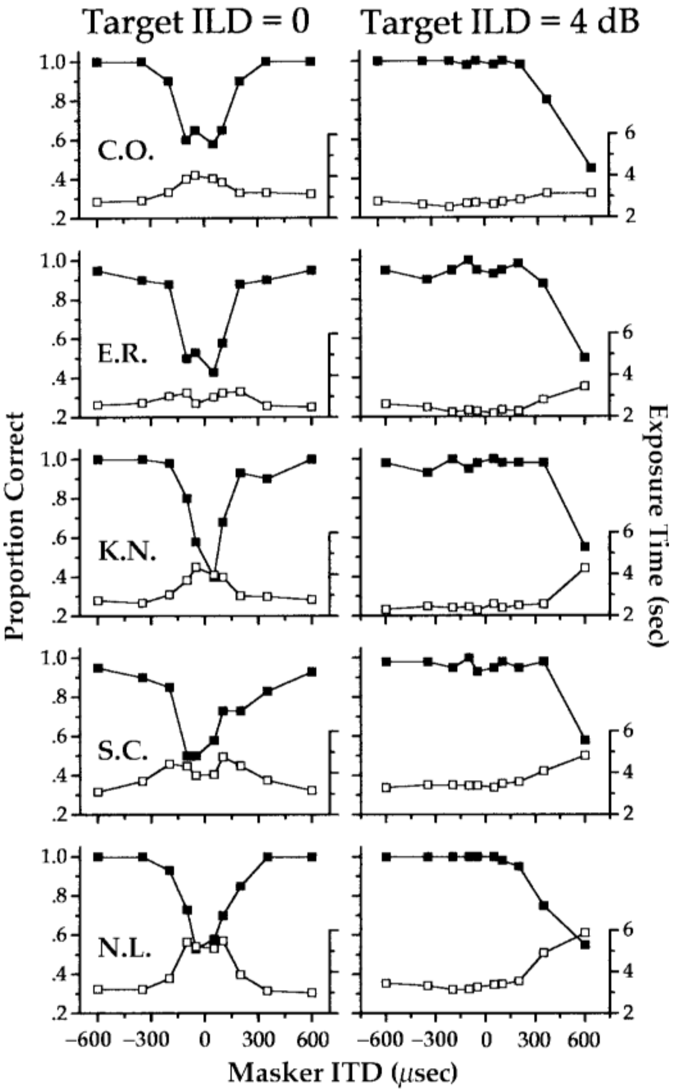

Figure 4. Results for Listeners 6–10. The left-hand panels show data from the target ILD = 0 condition, which for these listeners was experienced second. The right-hand panels display data from the target ILD = 4 dB condition. The ITD of the masker tones is plotted on the abscissa. Proportion correct (filled symbols) and exposure times (open symbols) are plotted on the ordinate; note the different scales.

rect data. When target and masker ITDs were similar, and thus when listeners performed less accurately, they chose to listen to a longer portion of the RMR sequence before responding. The results were not confounded by a speed–accuracy tradeoff, since this effect would have brought about the opposite pattern of results: Listeners would then have been more accurate on trials when they were slower in responding.

For several participants, a small aberration appeared at the tip of the tuning curve in the target ILD = 0 condition, such that maskers very near to the center, although closest of all to the targets, were slightly less damaging to performance than those a little farther away. It is not clear how much importance should be attached to this subtlety, but one might conjecture that some general processing advantage for sounds presented straight ahead outweighed even the disadvantage, from the standpoint of spatial segregation, of the near spatial coincidence of targets and maskers.

Of particular interest in relation to the matters raised in the introduction is what happened to the spatial tuning curves when the overall spatial position of the target was shifted to the right by introducing a target ILD at 4 dB, while the ITD remained at zero. To recall the specific pre-

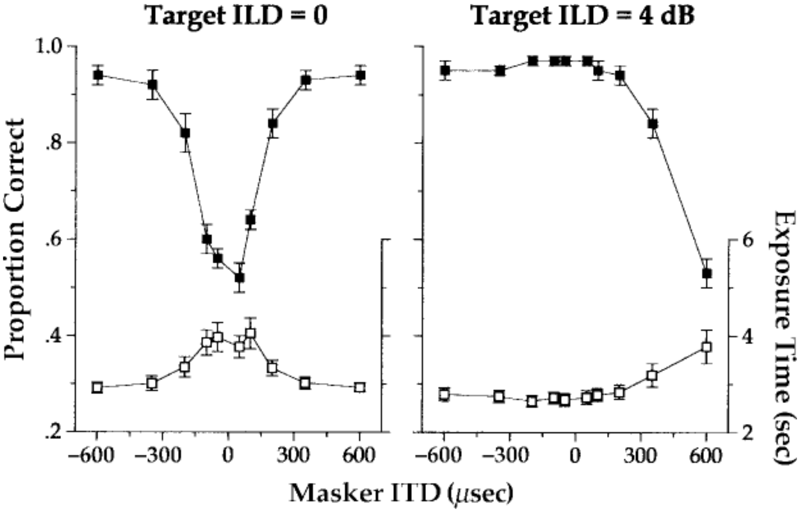

Figure 5. Summary of results across all 10 listeners. The left-hand panel shows data from the target ILD = 0 condition, and the right-hand panel displays data from the target ILD = 4 dB condition. The ITD of the masker tones is plotted on the abscissa. Proportion correct (filled symbols) and exposure times (open symbols) are plotted on the ordinate; note the different scales. Error bars correspond to one standard error of the mean.

dictions of the two hypotheses, illustrated graphically in Figure 1b: If attention operates early in the pathway at the level of ITDs, then the spatial tuning curve should remain unchanged from the target ILD = 0 dB condition. Since the target ITD remains at zero, the most distracting maskers should still be those with an ITD close to zero. In contrast, if attention operates late in the pathway, after the separately decoded ITD and ILD cues have been combined and integrated into an overall spatial percept, then the tuning curve should shift to the right. Since the overall position of the target has shifted, albeit by virtue of information from the ILD route, the most distracting maskers should be those whose ITD specifies a rightward lateral position equivalent to that resulting from an ILD of 4 dB.

The results are unequivocal: For every listener, the spatial tuning curve shifted to the right.[8] Once again, for those listeners who showed a systematic variation in RTs with masker ITD, the pattern was the reverse of that seen in the proportion correct data. These data argue strongly that at least within the constraints of the RMR paradigm, overall spatial position, and not ITD, is the substrate for auditory spatial attention. This result places attention relatively late in processing, in accord with the view that attention is subsequent to the integration of individual frequency components into holistic auditory objects (Darwin & Hukin, 1999).

REFERENCES

Bodden, M. (1996). Auditory demonstrations of a cocktail-party processor. *Acustica*, **82**, 356-357.

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound.* Cambridge, MA: MIT Press.

Bregman, A. S., & Ahad, P. A. (1996). *Demonstrations of auditory scene analysis* [Audio CD with notes]. Cambridge, MA: MIT Press.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon.

Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *Journal of the Acoustical Society of America*, **83**, 1508-1516.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 115-127.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.

Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *Journal of the Acoustical Society of America*, **98**, 785-797.

Darwin, C. J., & Hukin, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *Journal of the Acoustical Society of America*, **102**, 2316-2324.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 617-629.

Deatherage, B. H., & Hirsh, I. J. (1959). Auditory localization of clicks. *Journal of the Acoustical Society of America*, **31**, 486-492.

Drennan, W., Gatehouse, S., & Lever, C. (2003). Perceptual segregation of competing speech sounds: The role of spatial location. *Journal of the Acoustical Society of America*, **114**, 2178-2189.

Driver, J. (1996, May 2). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, **381**, 66-68.

Gilliom, J. D., & Sorkin, R. D. (1972). Discrimination of interaural time and intensity. *Journal of the Acoustical Society of America*, **52**, 1635-1644.

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. *Journal of Neurophysiology*, **32**, 613-636.

Griffiths, T. D., Elliott, C., Coulthard, A., Cartlidge, N. E., & Green, G. C. (1998). A distinct low-level mechanism for interaural timing analysis in human hearing. *NeuroReport*, **9**, 3383-3386.

Hafter, E. R., & Jeffress, L. A. (1968). Two-image lateralization of tones and clicks. *Journal of the Acoustical Society of America*, **44**, 563-569.

Hill, N. I., & Darwin, C. J. (1996). Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning. *Journal of the Acoustical Society of America*, **100**, 2352-2364.

James, W. (1890). *The principles of psychology.* New York: Holt.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative & Physiological Psychology*, **41**, 35-39.

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 nonspeech auditory patterns. *Journal of the Acoustical Society of America*, **104**, 422-431.

Klumpp, R. G., & Eady, H. R. (1956). Some measurements of interaural time difference thresholds. *Journal of the Acoustical Society of America*, **28**, 859-860.

Mills, A. W. (1958). On the minimum audible angle. *Journal of the Acoustical Society of America*, **30**, 237-246.

Mondor, T. A., & Amirault, K. J. (1998). Effect of same- and different-modality spatial cues on auditory and visual target identification. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 745-755.

Mondor, T. A., & Breau, L. M. (1999). Facilitative and inhibitory effects of location and frequency cues: Evidence of a modulation in perceptual sensitivity. *Perception & Psychophysics*, **61**, 438-444.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 387-409.

Moore, B. C. J. (1990). Comodulation masking release: Spectro-temporal pattern analysis in hearing. *British Journal of Audiology*, **24**, 131-137.

Olsen, J. F., Knudsen, E. I., & Esterly, S. D. (1990). Neural maps of interaural time and intensity differences in the optic tectum of the barn owl. *Hearing Research*, **47**, 235-256.

Perrott, D. R. (1984). Concurrent minimum audible angle: A reexamination of the concept of auditory spatial acuity. *Journal of the Acoustical Society of America*, **75**, 1201-1206.

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, **400**, 162-166.

Sach, A. J., Hill, N. I., & Bailey, P. J. (2000). Auditory spatial attention using interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 717-729.

Scharf, B. (1998). Auditory attention: The psychoacoustical approach. In H. Pashler (Ed.), *Attention.* Hove, U.K.: Psychology Press.

Schröger, E., Tervaniemi, M., Winkler, I., Wolff, C., & Näätänen, R. (1997). Processing of interaural cues used for auditory lateralization as revealed by the mismatch negativity. In A. Schick & M. Klatte (Eds.), *Contributions to psychological acoustics. Results of the seventh Oldenburg symposium on psychological acoustics.* Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg.

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 555-574.

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology*, **36A**, 51-74.

Teder, W., & Näätänen, R. (1994). Event-related potentials demonstrate a narrow focus of auditory spatial attention. *NeuroReport*, **5**, 709-711.

Turgeon, M. (1999). *Cross-spectral grouping using the paradigm of rhythmic masking release.* Unpublished doctoral dissertation, McGill University.

Turgeon, M., Bregman, A. S., & Ahad, P. A. (2002). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *Journal of the Acoustical Society of America*, **111**, 1819-1831.

Yin, T. C. T., & Chan, J. C. K. (1990). Interaural time sensitivity in the medial superior olive of the cat. *Journal of Neurophysiology*, **64**, 465-488.

NOTES

1. Significant interaural level differences will occur for laterally positioned sounds of any frequency when the sound source is close to the listener's head, simply as a result of the inverse square law.

2. The finding that the buildup of sequential grouping of sounds presented to one ear is reduced by attention to a competing task in the contralateral ear (Carlyon, Cusack, Foxton, & Robertson, 2001) invites some reappraisal of the simple hypothesis that stream segregation always precedes attention, which then selects among already formed streams.

3. For example, a conceivable explanation of reaction time effects (if not threshold changes) when all that must be accounted for is a simple difference between attended and probe trials is that reactions are slowed by the "surprise" caused by relatively infrequent probe trials.

4. Tones in which ITD and ILD cues are set in *opposition*, where one cue indicates a leftward and the other a rightward lateral position, have been used to demonstrate an ITD/ILD trading relationship, so that appropriate settings of the two opposed cues can give rise to a centered perceptual image. In some circumstances, the opposed lateralization cues do not trade completely and give rise to separate "time" and "intensity" images (Hafter & Jeffress, 1968). In the present experiment, ILDs and ITDs are never opposed: The target tones with ILD $= 4$ dB and masker tones with nonzero ITD have, respectively, ITDs and ILDs set to zero. Such tones give rise to unambiguous perceptual images at lateral positions determined by whichever of the interaural difference cues is set to be nonzero.

5. The MCF software was made available by kind permission of the copyright owner, Pierre Ahad. Details from digivox@hotmail.com.

6. The *equivalent rectangular bandwidth* (ERB) of a filter, often used as a measure of the auditory critical bandwidth, is the bandwidth of a rectangular filter that has the same peak transmission as that filter and that defined for the frequency domain, passes the same total power for a white noise input. The ERB-like measure described in the text provides an estimate of the width of the spatial tuning curve that is broadly similar to the width of the curve at a level of performance above the minimum corresponding to a factor of $\sqrt{2}$ of the total range of performance—a bandwidth estimate roughly analogous to the "half-power" bandwidth.

7. Of course the sounds used in our experiment were not simultaneous, but the target and maskers are interleaved in time rather as if they were concurrent cocktail party sources.

8. The range of masker ITDs used and the choice of ILD in the target ILD $= 4$ dB condition limit the data in that condition to tracing out only part of the tuning curve. The fact that for most participants, performance in the target ILD $= 4$ dB condition was close to chance at masker ITD $= +600$ μsec might be taken to suggest that the data show most of the left skirt of the spatial tuning curve at target ILD $= 4$ dB.

(Manuscript received August 27, 2003;
revision accepted for publication March 3, 2004.)