

Perceptual Asymmetries in Audition

Rhodri Cusack and Robert P. Carlyon
Medical Research Council Cognition and Brain Sciences Unit

Visual feature extraction has been investigated using search experiments. Targets that contain a feature not present in the distractors are easier to detect than if they do not, leading to search asymmetries. If sounds are decomposed into features in the auditory system, there might be asymmetries in analogous tasks. Six experiments investigating this are described. Strong asymmetries were identified, with frequency-modulated targets easier to detect among pure-tone distractors than vice versa and longer sounds easier to select from short distractors than the reverse. It is demonstrated that this asymmetry is not a result of peripheral limitations. In contrast, no asymmetries were observed between high- and low-frequency tones or between short 3-tone sequences differing only in their temporal structure. The results are discussed with reference to models of perceptual grouping and attention, the applicability of analogies between vision and audition, and possible physiological correlates. The paradigm provides a new way in which to investigate auditory feature extraction.

The sound arriving at the ears is often a mixture from different sources, and to interpret it, we usually have to select the part of the signal that comes from the source in which we are interested. This was coined the “cocktail party problem” by Cherry (1953). A valuable aid in its solution is perceptual grouping: Because of commonalities in natural sounds, a set of simple rules can help to group together parts of the auditory scene that come from the same source. For example, sounds that start simultaneously, or are harmonically related, are likely to have come from the same source and are grouped together. We can also group sounds over time; for example, those that are similar in frequency or arise from the same location in space are likely to be allocated to the same perceptual stream. Many such perceptual grouping processes have been well characterized (see Bregman, 1990; Darwin & Carlyon, 1995, for reviews).

Once sounds are grouped together, they can be accepted or rejected for further analysis as a whole, for example, when we attend to a high-pitched voice and ignore a low one. However, it will not always be possible for perceptual grouping processes to split the input into target and nontarget streams. In such cases, we have to use more actively our prior knowledge to select the target. These selective attention processes have received substantially less investigation than their visual counterparts; although initially discussed in the auditory domain

(e.g., Broadbent, 1958), the majority of subsequent work has focused on vision.

A classic laboratory task that has yielded much useful information about the processes underlying visual selective attention is *visual search*. A complex display, often a single static frame, is presented, and the participant is asked whether a specified target is present. Reaction time or accuracy on this visual search task is measured as a function of target type and the number and nature of the interfering distractors in the display. A well-established finding is that strong asymmetries occur. Although a particular target (e.g., a *Q*) may be easy to detect even when many distractors (e.g., *O*s) are present, the converse task (detection of an *O* among *Q*s) can be very difficult. These phenomena are particularly striking because the basic visual demands are similar, illustrating that the limitations are not sensory. Many permutations of distractors and targets have been tested, and several models have been proposed to account for the findings (e.g., Duncan & Humphreys, 1989; Triesman & Gelade, 1980; Treisman & Gormican, 1988).

One of the first models, *feature integration theory* (Triesman & Gelade, 1980), which is still extremely influential, asserted that early in visual processing, we form a number of separate feature maps. So, for example, we have a spatial array of straight-line detectors and a separate array of circle detectors. If the target uniquely contains a feature, then the entire array of detectors for this feature may be examined in parallel. This model fits empirical observations: For certain simple search tasks where the target contains a unique feature, the number of distractors in the display has little effect on performance (*pop-out*). However, if a *conjunction* of features (e.g., absence of a line and presence of a circle) must be searched for, then it is argued that a serial attentional process is necessary to bind the features of each object together, and the search is slower. Indeed, for such searches, it has been observed that the time taken is roughly proportional to the number of items in the display. Another characteristic of serial search observed in these cases is that the average time taken to say “absent” is twice as long as the time taken to say “present,”

Rhodri Cusack and Robert P. Carlyon, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom.

Some of the experiments described here were presented in summary form at the British Society of Audiology’s Short Papers Meeting, University of Essex, Colchester, United Kingdom, September 2000 (see Cusack & Carlyon, 2000, for abstract). We thank Mei-Chern Leow for help with some of the experiments and Michael Kubovy, Tim Green, and Andy Oxenham for their comments on a draft of this article.

Correspondence concerning this article should be addressed to Rhodri Cusack, Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom. E-mail: rhodri.cusack@mrc-cbu.cam.ac.uk

presumably because in the latter case, on average, only half the items in the display need to be searched before the target is found. Other related perceptual asymmetries that appear to result from the representation of sounds by feature detectors have also been identified. It can be easier to select a target from distractors when the target contains a greater activation on some particular dimension. So, for example, Beck (1974) has shown that detection of a longer line from among a set of short lines is easier than vice versa. Treisman and Gormican (1988) have demonstrated a similar effect using a reaction time measure. Such asymmetries in the detectability of changes to a single feature can be explained by applying Weber's law to feature detectors: The size of a just-detectable change in a feature is larger if there is more of that feature present overall (Treisman & Gormican, 1988). In a search task, there are more distractors present than targets, and these will dominate the pooled amount of activation in feature detectors. Hence, detection thresholds are lower (e.g., line-length discrimination better) when distractors have less of a feature (e.g., short lines) than when they have more of it (e.g., long lines).

Despite the appealing simplicity of the serial versus parallel search model, it has been argued that there is not a dichotomy between feature and conjunction searches, but a continuum between easy and difficult ones (e.g., Duncan & Humphreys, 1989). However, although the precise mechanisms underlying these asymmetries are still the subject of some debate, it remains true that the presence or absence of features can have a strong effect on performance. Further evidence that we extract features in vision comes from neurophysiological and neuroimaging studies. Different neural structures have been shown to be selectively responsive to a whole range of different features, such as orientation, size, color, or direction of movement (F. W. Campbell & Maffei, 1970; Cowey, 1985; Zeki, 1978, 1981).

In the auditory domain, several authors have also argued for feature extraction processes (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Howard & Ballas, 1981; Tou, 1981). The most efficient way to recognize a noisy stimulus, they argue, is to extract its essential characteristics and then make a judgment on these. There is also neurophysiological evidence for feature extraction. Neurons selectively responding to frequency modulation (FM) have been identified in the rat (Rees & Møller, 1983), cat (see Ehret, 1997, for review), and monkey (Bieser, 1998), and neurons selective for amplitude modulation have been identified in the frog (Rose & Capranica, 1983), bird (Muller & Leppelsack, 1985), and rat (Rees & Møller, 1983). Rauschecker (1998) has proposed that particular regions in the lateral belt of auditory cortex in primates are tuned by bandwidth. Several studies have found duration-sensitive neurons in the bat (Casseday, Ehrlich, & Covey, 1994; Fuzessery & Hall, 1999; Galazyuk & Feng, 1997; Ma & Suga, 2001; Pinheiro, Wu, & Jen, 1991), mouse (Brand, Urban, & Grothe, 2000; Xia, Qi, & Shen, 2000), chinchilla (Chen, 1998), and cat (He, Hashikawa, Ojima, & Kinouchi, 1997). Some electrophysiological studies using scalp recordings in mismatch negativity (MMN) paradigms have shown that preattentive processing of sound involves the extraction of features (e.g., Deacon, Nousak, Pilotti, Ritter, & Yang, 1998; Paavilainen, Simola, Jaramillo, Näätänen, & Winkler, 2001). Given this stage of feature extraction, we might expect to see

perceptual asymmetries in audition. The aim of the current study was to develop tasks analogous to those used in visual search to investigate this.

Experiment 1

Method

As discussed in the introduction, there is good evidence that the auditory system contains neurons that act as FM detectors. In Experiments 1–3, we use FM for this reason and because it has the advantage that it can be added to a pure tone, leaving the sound's duration, center frequency, and approximate spectral distribution unchanged. Search experiments in vision have most commonly used a large number of items, but small search sets have been successfully used to investigate asymmetries (e.g., four items in studies by Beck, 1974, and Williams & Julesz, 1992). Because there was a procedure analogous to these small search set experiments in common use in psychoacoustics, we used it in our initial studies. We tested whether listeners found it easier to identify a single FM tone among pure tones than vice versa, using a two-alternative three-interval forced-choice (2A-3IFC) task. In later experiments, we used stimuli with a greater number of items that were more directly comparable to those commonly used in visual search experiments. Four listeners in the age range of 20 to 30 years, all of whom reported normal hearing, took part in Experiment 1.

On each trial, we presented three sounds sequentially (each 500 ms in duration, separated by 500-ms gaps, with a center frequency of 500 Hz). In Condition 1, two of these sounds were pure tones, and one (either the second or the third) was an FM tone (see Figure 1). In Condition 2, two of the sounds were FM tones, and one (either the second or the third) was a pure tone. In Condition 1, listeners were asked to identify the "warble," and in Condition 2, listeners were asked to identify the "steady tone."

Because it is more difficult to interpret reaction times when the stimuli are spread over time, we used an accuracy measure. Note that evidence from visual experiments suggests that perceptual asymmetries can be revealed in this way (e.g., Beck, 1974). To minimize floor and ceiling effects, we conducted a pretest in which the FM threshold of each participant was measured. The pretest used a one-interval task in which a single sound (500 ms long, 500 Hz center frequency) was presented on each trial. On half of the trials, it was a tone frequency modulated at 2 Hz, and on half of the trials, it was an unmodulated tone. Listeners were asked to respond "steady" or "warble." A two-up one-down adaptive procedure (Levitt, 1971) was used, with an initial peak-to-peak FM depth of 1.5 semitones and a step size of 0.05 semitones. The procedure was terminated after 12 turn points, and the average of the last 6 were taken as the threshold. Each participant did a single practice block, and then two main blocks. Their thresholds ($M = 0.095$, $SD = 0.040$ semitones) were then used in the main experiment.

For the main experiment, we presented two blocks of 64 trials of each condition in an ABBA order, counterbalanced over participants. Before the first block of each condition, 8 practice trials with a large FM depth (1 semitone) were used to demonstrate the task. The sounds were presented diotically; each tone had a level of 60 dB SPL. In all experiments,

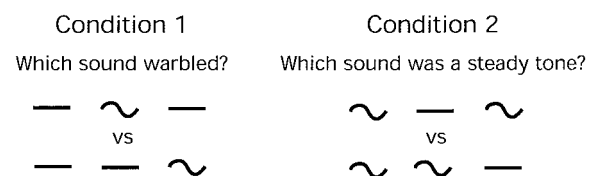


Figure 1. Stimuli used in Experiment 1.

responses were recorded by keypress, and no feedback was given. All stimuli were generated using a PCMCIA soundcard in a portable PC, at a sample rate of 22050 Hz, and presented via Sennheiser HD-414 headphones in a quiet room. To avoid any effects of response bias, performance is described using the signal-detection measure d' (Macmillan & Creelman, 1991).

Results

Figure 2 illustrates the results. Listeners were much worse at identifying the interval containing the odd steady tone (Condition 2) than identifying the odd warble (Condition 1). A one-way repeated measures analysis of variance (ANOVA) with the Huynh-Feldt sphericity correction on the d' scores showed a strong effect, $F(1, 3) = 299.6, p < .001$. The c measure (Macmillan & Creelman, 1991, p. 33) was used to examine any response bias in this experiment and Experiments 3–6. Most comparisons showed no significant change in bias; two relating to dichotic conditions in Experiments 5 and 6 were significant and are discussed in their respective *Results* sections.

Discussion

The results demonstrate a perceptual asymmetry. Although other explanations will be discussed in a later section, a candidate explanation is that, in analogy with findings in visual selective attention and in concordance with cognitive models of sound processing and neurophysiological evidence, FM is coded as a feature. It is easier to select a sound containing FM than to identify the presence of a tone and the absence of FM. This is as would be predicted from a model similar to that used by Treisman & Gormican (1988). When the pooled activation in the feature detectors is greater (i.e., Condition 2) than from Weber's law, detection performance is expected to be worse. Here, we are pooling activation across time, as one would expect when this is the dimension that must be searched for the target. Note that our results are not as predicted by signal-detection theory and show a strong asymmetry in a procedure (2A-3IFC) that most auditory

scientists assume to reflect purely sensory processes (Hanna, 1992).

Experiment 2

The asymmetry observed in Experiment 1 was consistent with the idea that targets are easy to select when they contain an extra feature. However, we have not established that asymmetries are *not* observed in situations in which we would *not* expect asymmetric activation of a set of feature detectors. Hence, in this experiment, we used a similar task (although with a slightly different measurement procedure), but in addition to conditions in which targets were distinguished by the presence or absence of frequency modulation, we included conditions in which all sounds were steady tones and targets were distinguished by being lower or higher in frequency. We would not expect there to be an asymmetry in the degree to which feature detectors are activated by lower and higher frequency tones. It is well established that there is channeling by frequency at many stages of the auditory system, and so whether the target is lower or higher in frequency than the standards, it will excite a different (although possibly overlapping) population of neurons. There will also be differences in the temporal pattern of neural firing to sounds of different frequencies, but there is no evidence for an asymmetry between total firing rates to lower and higher frequency sounds. Feature detection should be symmetrical, and so we should not expect to see any effect of the sign of the mean frequency difference that distinguishes the target.

Method

Eight listeners who reported normal hearing were tested in four conditions. None had participated in the previous experiment. A 2A-3IFC task was used with an adaptive procedure. On each trial, listeners were presented with three sounds that were each 500 ms in duration and separated by 500-ms gaps. Two of the sounds were standards, and one (either the second or the third) was a target. Conditions 1a and 1b required FM detection, and Conditions 2a and 2b required frequency discrimination. In Condition 1a, the standards were steady tones, and the targets were tones frequency modulated at a rate of 2 Hz; and vice versa in Condition 1b. In Condition 2a, the standards were lower in frequency and the target higher and vice versa in Condition 2b. The order of presentation of the a and b conditions was counterbalanced. In Conditions 1a and 1b, the depth of frequency modulation was adaptively varied, whereas in Conditions 2a and 2b, the frequency difference between standard and target was varied. For all conditions, the adaptive parameter was varied using a two-up one-down procedure, with a starting value of 0.3125 semitones and a step size of 0.025 semitones. The first four trials of each run were for practice, and the adaptive variation of depth did not begin until after these. The parameter was constrained to have a minimum of 0.0125 semitones. An adaptive run was terminated after 12 reversals, and the mean of the final 6 was taken as the threshold. Two adaptive runs of each condition were performed. The sounds were presented diotically; each sound had a level of 75 dB SPL.

Results

Figure 3 shows the thresholds measured. As in Experiment 1, there was an asymmetry between Conditions 1a and 1b, with thresholds higher when a steady target had to be identified among two frequency-modulated tones than vice versa (repeated measures ANOVA; the Huynh-Feldt correction), $F(1, 7) = 9.783, p < .02$. However, in the frequency-discrimination task, there was no evi-

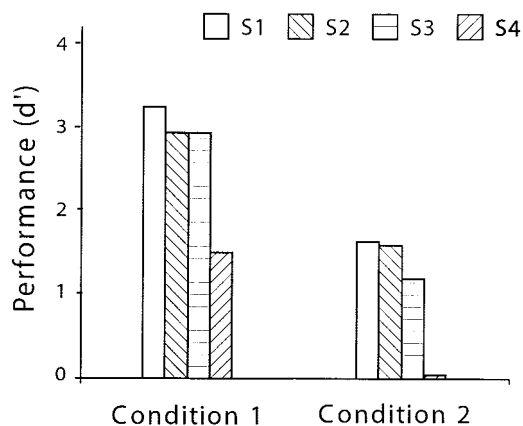


Figure 2. Performance of 4 participants (S1–S4) in Experiment 1, when a single frequency modulation (FM) tone had to be selected from two pure tones (Condition 1) and when a single pure tone had to be selected from two FM tones (Condition 2).

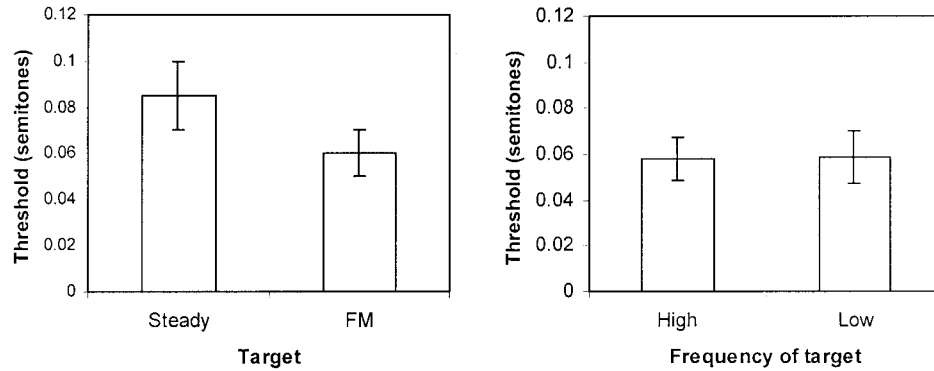


Figure 3. Thresholds for frequency modulation (FM) detection (left) and frequency discrimination (right) in Experiment 2. Means and standard errors of the means for 8 participants are shown.

dence of an asymmetry dependent on which of the higher or lower frequency tone was the target, $F(1, 7) = 0.032$, *ns*. A combined ANOVA with the Huynh–Feldt sphericity correction showed a Stimulus Type \times Asymmetry interaction, $F(1, 7) = 7.55$, $p < .05$.

Discussion

An asymmetry was observed using an adaptive procedure in an FM detection task, but not when a very similar procedure was used to measure frequency-discrimination thresholds. This is consistent with our hypothesis that there are specific feature detectors for the presence of FM but symmetrical detectors for frequency changes.

There is an important difference between these two experiments and classic visual search. In the visual tasks, the feature to be detected is often highly suprathreshold. There remains an interesting question: Can perceptual asymmetries still be seen with suprathreshold differences distinguishing auditory targets?

Experiment 3

Method

Four listeners who reported normal hearing were tested. None had participated in the previous experiments. We again used FM as the characteristic distinguishing the targets. Because we made the task easier by increasing the FM depth, if we were to again use an accuracy measure, we needed to make the task more difficult in some other way. To do this, we increased the number of distractors, using a design more directly analogous to those classically used in visual search. As was noted by Kubovy (1981) and Kubovy and van Valkenburg (2001), when making analogies between vision and audition, there are several different ways in which stimulus dimensions may be mapped onto each other. We translated the spatial dimensions into time and frequency because we felt that these dimensions rather than spatial ones were the most important for distinguishing auditory events. This is discussed further in the General Discussion.

Schematics of stimuli in each condition are shown in Figure 4. On each trial, 4, 8, 16, or 32 tones of duration 250 ms were randomly distributed over a 2-s window. They were also randomly distributed in frequency (262–4192 Hz) with a logarithmic distribution, and subject to the constraint that no two simultaneous tones could come closer than one third of an octave to each other. In Condition 1, the distractors were pure tones. On half of the trials, one of the tones, a target, was modulated at a depth of 5% (0.84 semitones) and a rate of 8 Hz. In Condition 2, the distractors were FM

tones, and on half the trials, one of the tones was an unmodulated target. The listener's task was to determine whether the target sound was present.

Two blocks of 128 trials of each condition were presented in an AB order counterbalanced over participants. There were equal numbers of trials with each number of distractors, which were randomly intermixed. Before the first block of each condition, there was a practice block of 64 trials. A d' measure was used to remove the effects of response bias. The sounds were presented diotically; each tone had a level of 72 dB SPL.

Results

The results of Experiment 3 are shown in Figure 5. Listeners were much better at detecting the target when it was an FM tone than when it was a pure tone; a two-way repeated measures ANOVA with Huynh–Feldt sphericity correction showed a main

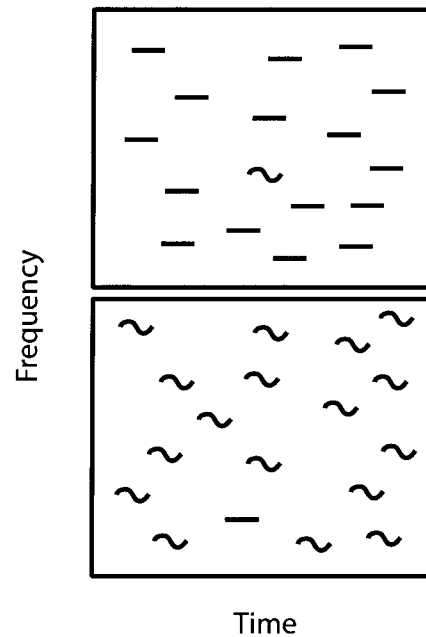


Figure 4. A schematic of example stimuli from Experiment 3: Condition 1 (top) and Condition 2 (bottom).

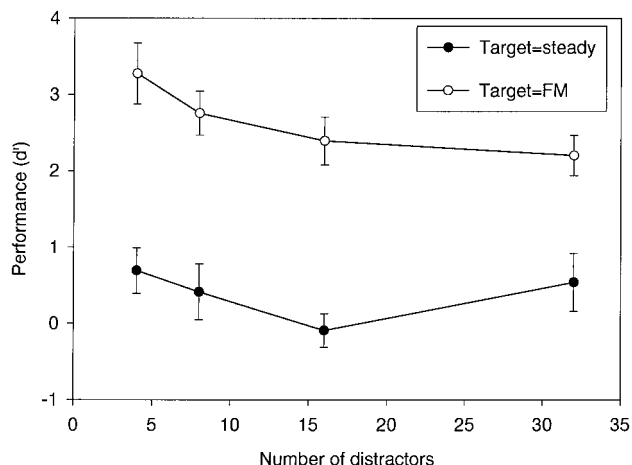


Figure 5. Mean performance (± 1 SEM) of the 4 participants in Experiment 3. FM = frequency modulation.

effect of condition, $F(1, 3) = 18.1$, $p < .05$; a main effect of number of distractors, $F(1, 3) = 14.47$, $p < .005$; but no Condition \times Number of Distractors interaction, $F(3, 9) = 0.60$, *ns*.

Discussion

It was much easier to detect the FM tone in a background of pure tones than vice versa. These results are what would be expected if there were a perceptual asymmetry, which remains even when the difference between an isolated target and distractor are clearly suprathreshold.

Although there was an asymmetry, in a true parallel search condition, we might expect performance to be unaffected by an increasing number of distractors. In our experiment, there was a decline in performance with increasing numbers of distractors in both conditions, and there was not a Task \times Distractor Number interaction. This suggests that we are observing a perceptual asymmetry and not pure pop-out in the easier condition, which is an example of the continuous rather than dichotomous nature of search tasks from difficult to easy, as demonstrated by Duncan and Humphreys (1989). When the target was unmodulated, there was a trend toward an improvement in performance as the number of distractors was increased from 16 to 32, $t(3) = 3.64$, $p < .05$, not corrected for multiple comparisons. This was perhaps the result of an effect previously observed in informational masking studies where a larger number of distractors can lead to less masking as they become more homogeneous from trial to trial (Oh & Lufti, 1998; see General Discussion).

In Experiment 3, the minimum separation between the target and any simultaneously occurring distractor was equal to one third of an octave (26%). Because the targets and distractors had equal levels, this meant that in the absence of attentional limitations, the target would always be easily detectable (Glasberg & Moore, 1990). However, it is possible that there may have been some partial masking of the target. This does not affect our conclusions as long as the extent of such partial masking did not differ between steady and modulated maskers. One possible reason why this may not be true is that the frequency spectrum of FM tones contains sidebands separated from the carrier at integer multiples of the

modulation rate, causing the overall bandwidth of the stimulus to be broader. The distribution of amplitudes among the sidebands depends critically on the modulation depth, and for the modest FM depth used here, the distribution would have been strongly dominated by those components closest to the carrier frequency. Because the overall extent of these sidebands is substantially smaller than the bandwidths of peripheral auditory filters (Glasberg & Moore, 1990), we would expect the peripheral excitation patterns (and hence the masking potential) of our steady and FM tones to have been very similar. However, as discussed in the next section, there would have been a very slightly wider spread of neural excitation for the FM tones than for the steady tones. Hence it is just possible that an FM distractor would have produced more partial masking of a steady tone than vice versa. The next experiment was designed to investigate this possibility.

Experiment 4

Method

The stimuli in Experiment 4 were identical to those in Experiment 3, except for the following changes. There were four conditions: In Conditions 1a and 1b, the target was an unmodulated tone, and the distractors were FM tones, and in Conditions 2a and 2b, the converse was true. To control for any possible differential overlap in the neural excitation in the ear, we calculated the excitation patterns of the sounds using Glasberg and Moore's (1990) model, which is based on the results of notched-noise masking experiments. Figure 6 shows the excitation patterns for a steady (dotted line) and FM (heavy solid line) tone having a carrier frequency at the geometric mean of the range for our stimuli (1024 Hz). It can be seen that the excitation pattern of the FM tone is indeed a little broader than that of the steady tone. However, if the carrier frequency of the FM tone is reduced by 1 semitone, then it produces less excitation along all points on the upper skirt of its excitation pattern (left-hand solid line) than does the steady tone. Similarly, if the carrier frequency of the FM tone is increased by 1 semitone, it produces less excitation on its lower skirt than is seen for the steady tone. This means that if we constrain the FM distractors to be 1 semitone further away from the targets than are the steady distractors, then a peripheral interaction hypothesis would predict that the latter would lead to worse performance. In contrast, if our original feature-detection hypoth-

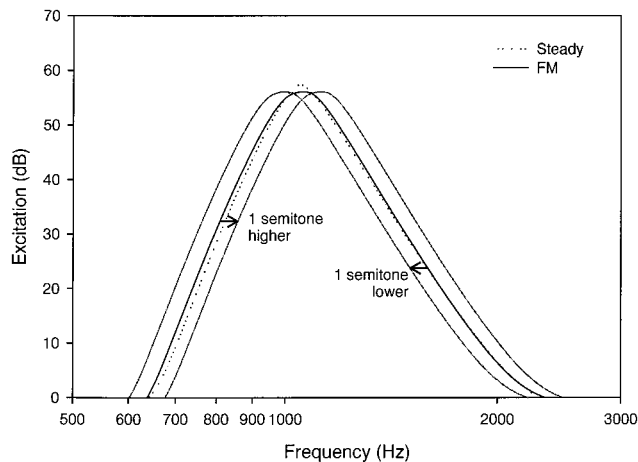


Figure 6. Neural response to the stimuli used in Experiment 3 derived from Glasberg and Moore's (1990) model of the peripheral auditory system. FM = frequency modulation.

esis was correct, then performance would be worse with FM distractors and steady targets.

To test this, the minimum separation between targets and distractors was 5.5 semitones in Condition 1b, but 6.5 semitones in Condition 2b. For Conditions 1a and 2a, the separation was equal at 6 semitones. This design allowed us to probe the effect of peripheral masking in two ways. First, if the results of Experiment 3 were due to differential overlap of excitation, we might expect the difference between Conditions 1a and 2a to be reduced in Experiment 4, because the minimum separation had been increased from one-third octave (4 semitones) to half an octave (6 semitones). Furthermore, the difference would disappear entirely or reverse between Conditions 1b and 2b.

Unlike in Experiment 3, in this experiment, we fixed the number of distractors at 16. Two blocks were presented, one of which comprised 64 trials of each of Conditions 1a and 1b and the other 64 trials each of Conditions 2a and 2b. Within a block, the order of presentation was randomized, whereas the order of presentation of the two blocks was counterbalanced across participants. Eight participants who reported normal hearing were tested. Seven of the 8 had participated in Experiment 2. Each tone had a level of 78 dB SPL.

Results

Figure 7 shows performance in the four conditions. Even with a minimum frequency separation of half an octave, there is a strong asymmetry (repeated measures with Huynh–Feldt correction: Condition 1a vs. Condition 2a), $F(1, 7) = 25.8, p < .002$. Furthermore, with the differential minimum separations, there remained a very strong effect (Condition 1b vs. Condition 2b), $F(1, 7) = 45.7, p < .001$, showing that differential overlap of excitation was not responsible for the results. An additional confirmation was obtained by testing whether the size of the asymmetry reduced in this experiment relative to Experiment 2. We entered the data from Conditions 1a and 2a in this experiment and the data from the 16-distractor condition of Experiment 2 into a single ANOVA. A between-subjects model was used because the two experiments used different participants. As the frequency spacing between sounds was larger in this experiment than in Experiment 2, if the asymmetry was an effect of peripheral masking, it would be

expected to be smaller. However, no such effect was found, $F(1, 20) = 1.04, ns$.

Discussion

Experiments 1–4 demonstrate a perceptual asymmetry. However, we have not investigated the degree to which the effect will generalize to other features. There is a possible alternative explanation that might account for some of the results and might be specific to FM. It is possible that our results might be affected by differences in the degree of adaptation to FM in the two conditions. In both experiments, in the condition in which performance was worse, there was a greater number of FM tones. In previous experiments with FM, it has been shown that under particular circumstances, adaptation can occur (e.g., Gardner & Wilson, 1979; Wakefield & Viemeister, 1984). We think that it is unlikely that adaptation was responsible for the effects we observed, because the conditions in those experiments are usually very different to those we have used. For example, Gardner and Wilson generated adaptation to frequency sweeps (*linear* FM) by alternating a burst of 20 large frequency sweeps (around 80 times threshold) with test stimuli. Also, given the extended time course of adaptation, which is of the order of many minutes (Tansley & Suffield, 1983), we would expect an effect of position within a block, with greater adaptation, and hence worse performance, in later trials. Figure 8 shows the mean performance in Experiment 3 for the 4 participants across time. The experiment was divided into subblocks, each containing 16 trials, and within each subblock, we calculated d' s and took means across participants. No such trend toward worse performance in later trials was seen, and there was a large asymmetry between performance in the very first trials.

As a further test of our conclusion that perceptual asymmetries can be observed in the absence of adaptation to FM, we repeated the experiment with stimuli that have a distinguishing feature for which there is no evidence of adaptation.

Experiment 5

Method

In this experiment, we used a similar procedure to Experiment 3. However, instead of distinguishing the target by the presence or absence of FM, it was different in its duration. Several studies have demonstrated the presence of duration-specific neurons (see summary in introduction and *Discussion* of this experiment). Thus, duration seems a possible candidate for something coded in a population of feature detectors and something for which we might observe perceptual asymmetries.

On each trial, the listener's task was again to determine whether a target was present or absent. There were three conditions. Conditions 1 and 2 were analogous to those in Experiment 3 (see Figure 9). In Condition 1, the distractors were 100 ms long. On half of the trials, a longer target was present (300-ms tone). In Condition 2, the distractors were 300 ms long, and the target duration was 100 ms. The tones were again randomly distributed in a time window that was 2 s long. Again they were also randomly distributed in frequency (262–4192 Hz) with a logarithmic distribution, and subject to the constraint that no simultaneous tones could come closer than one third of an octave. To equate the total distractor energy in the two conditions, the number of 100-ms distractors (3, 12, 21, or 30) was three times higher than the number of 300-ms distractors (1, 4, 7, or 10). Note that if it is the number of distracting sounds per se that

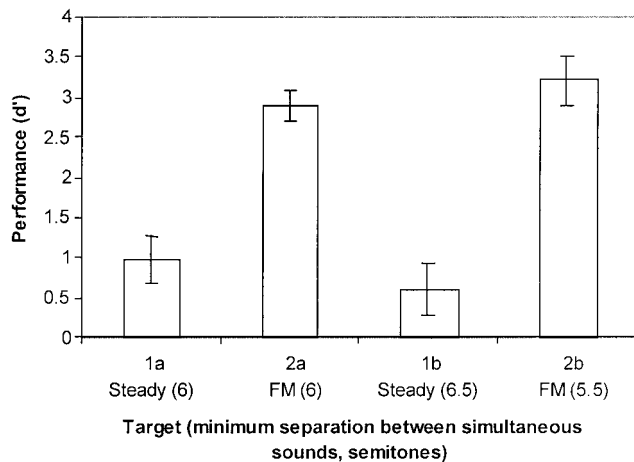


Figure 7. Performance in Experiment 4. Error bars represent one standard error of the mean FM = frequency modulation.

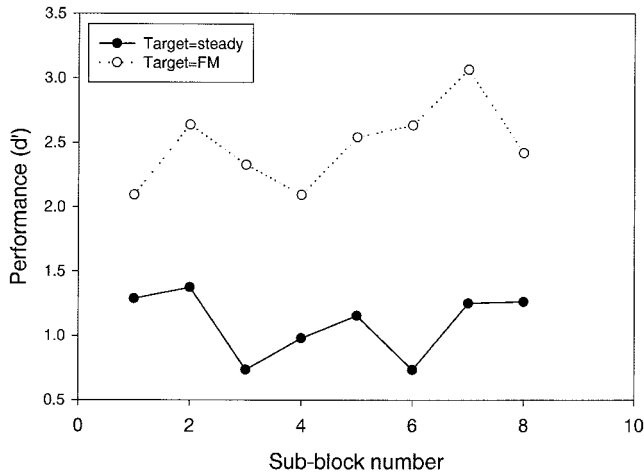


Figure 8. Time course of performance in Experiment 3. Mean performance for the 4 participants within each block. The subblocks comprise 16 trials each. FM = frequency modulation.

modulates the search difficulty, worse performance would be expected in Condition 1 than in Condition 2.

Although we equated the amount of distractor energy in the two conditions and reduced peripheral masking by constraining the stimuli so that two simultaneous sounds never came closer than one third of an octave, we felt it would be useful to control explicitly for low-level peripheral effects. To do this, we compared Conditions 1 and 2, in which target and distractors were presented only to the left ear, with Condition 3, which was identical to Condition 2 except that the distractors were additionally presented to the right ear. This diotic presentation of the maskers caused them to be heard in the center of the head, whereas the target remained in the left ear. Note that when the target was absent, one of the distractors did not have a copy in the right ear, to prevent participants from performing the task simply by hearing one component in a different location to the rest; rather, participants still had to perform a duration discrimination. If some form of peripheral masking was the limiting factor in the experiment, then as the sounds in Conditions 2 and 3 were identical in the left ear, performance in these two conditions would be similar. However, if more central limitations were the cause, then the difference in the binaural configuration of the

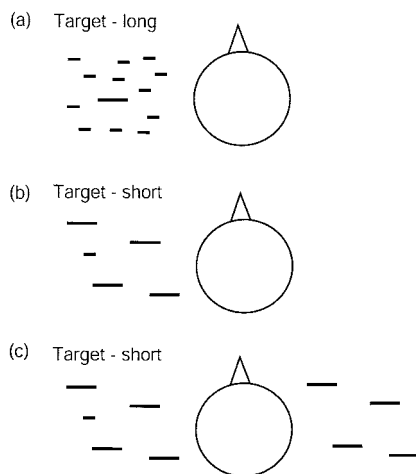


Figure 9. A schematic of example stimuli in Experiment 5.

target and distractors in Condition 3 would reduce the attentional load and performance would be improved.

Two blocks of 64 trials of each condition were presented in an ABCBBA order counterbalanced over participants. The trials with each of the different numbers of distractors were randomly intermixed. Before the first block of each condition, there were two practice blocks. The first comprised 16 trials in which only one or three distractors were present, and the second comprised 64 trials similar to those used in the main experiment. Six listeners who reported normal hearing participated. None had participated in the previous experiments. Each tone had a level of 73 dB SPL.

Results

Figure 10 shows performance as a function of the total distractor duration. Note that there were three times as many short distractors as there were long ones. All three conditions were entered into a repeated measures ANOVA with the Huynh-Feldt sphericity correction. There were significant main effects of condition, $F(2, 10) = 10.92, p < .005$, and total distractor duration, $F(3, 15) = 39.5, p < .001$, as well as a Condition \times Total Distractor Duration interaction, $F(6, 30) = 7.85, p < .001$. The ANOVA was also repeated on the data from just Conditions 1 and 2 to test for a difference between the short- and long-target monotic conditions. This again showed main effects of total distractor duration, $F(3, 15) = 33.8, p < .001$, and condition, $F(1, 5) = 9.18, p < .05$, and a Condition \times Total Distractor Duration interaction, $F(3, 15) = 10.09, p < .001$. The criterion measure c was calculated for each condition and participant to investigate differences in response biases between conditions. There were no main effects, but there was a Condition \times Total Distractor Duration interaction, $F(6, 30) = 4.74, p < .005$. Examination of the mean c scores averaged over participants suggested that when there were a large number of distractors, participants were less likely to say the target was present in the monotic than the dichotic condition. This suggests that as the task became more difficult, they adopted a more conservative criterion.

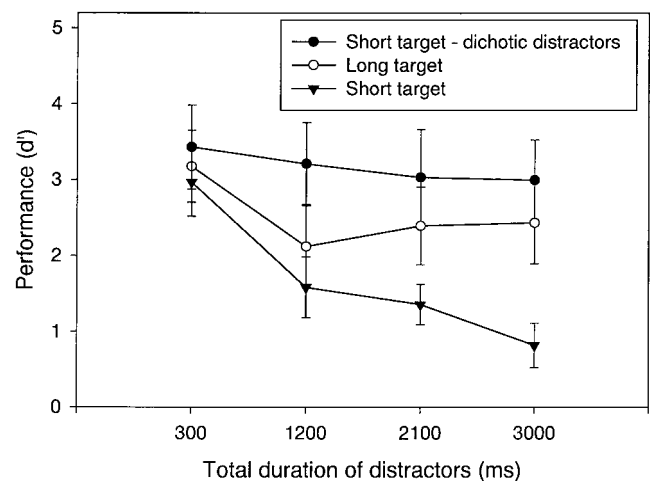


Figure 10. Mean performance (± 1 SEM) of 4 participants in Experiment 5.

Discussion

For monotic presentation, a strong perceptual asymmetry was observed, with a longer target embedded in many short distractors being much easier to detect than the converse. This result is analogous to one obtained in vision by Beck (1974), in which longer lines were easier to detect in a background of shorter lines than the converse. The significant interaction between number of distractors and condition is in the same direction as would be expected if pop-out were occurring for long targets embedded in short distractors. The FM-specific explanation suggested for the results of Experiments 1–4 would not predict any effect in Experiment 5. The most parsimonious conclusion therefore is that the asymmetries observed have a common cause.

Additionally, when a copy of the distractors was added to the opposite ear, performance improved substantially, in keeping with a reduction in attentional load. We believe that the most likely reason for this was that the contralateral copy of the distractors caused them to be heard in a different spatial location to the target and that participants could therefore monitor the target location when performing the task. That performance was not worse in this condition than in Condition 1 (the long-target condition) and was substantially better than in Condition 2 (short-target monotic condition) confirms that we successfully controlled peripheral masking in Conditions 1 and 2 and that the limiting factor in Condition 2 was not the loss of information at an early stage in auditory processing.

The perceptual asymmetry observed in our experiments suggests that the 300-ms sounds excite some population of feature detectors to a greater extent than the 100-ms sounds. This is broadly consistent with the available electrophysiological data, even though this has been collected in a range of animals, often with stimuli very different from our own. He et al. (1997) recorded responses in the auditory cortex of the cat to noise bursts having durations in the range 20–500 ms. They found 59% of their duration-sensitive neurons were long-pass (i.e., responding more to longer sounds), and from the typical data given, these responded much more strongly (2–7 times increase in spike rates) to longer sounds (200–300 ms) than to 100-ms sounds. The remainder of the duration-sensitive neurons were short-pass and temporal band-pass units. These typically responded most strongly to very short sounds (e.g., 20–50 ms) and had responses that decreased only slightly (perhaps 10%–30%) between 100 ms and the longest duration for which data are given (200 ms). In the mouse, Brand et al. (2000) used various sounds of durations in the range 1–100 ms and recorded from the auditory midbrain. They found that the majority of neurons (69% and 73% in two experiments) were long-pass. The neurons that were short-pass or temporal band-pass had upper cutoff durations (20–70 ms) below that of our 100-ms stimuli and showed no response to the 100-ms sound. Xia et al. (2000) recorded responses in the inferior colliculus of mice to pure tones with durations in the range of 3 to 300 ms. They found that 49% of their duration-selective neurons were long-pass and all of the typical units shown demonstrated a strong increase (2–4 times) in response strength as the stimuli were increased in length from 100 ms to 300 ms. In contrast, the short- and band-pass neurons showed weak responses, and only a few were found that appeared to show a greater response (1–2.5 times) to the 100-ms tone than to the 300-ms tone. A number of studies in bats have shown

duration tuning, although these studies are perhaps less relevant, as the bat auditory system is specialized for processing echolocation sounds, which are often brief, and the stimuli used in the experiments were usually rapid FM sweeps. Although some of these bat studies have shown a greater number of neurons tuned to short than to long durations, the short-pass units responded only to very short sounds (<20 ms; Fuzessery & Hall, 1999; Galazyuk & Feng, 1997; Zhou & Jen, 2001). Furthermore, a study of the effect of sound intensity on duration tuning showed that as sounds are presented at levels well above threshold, long-pass neurons remain duration selective, but short-pass neurons start to respond to all sounds (Zhou & Jen, 2001). In summary, our data are broadly consistent with data from a range of animals, even though they were collected with somewhat different stimuli. If the human auditory system does contain a similar set of neurons tuned to very short durations, then we can make the extremely interesting prediction that perceptual asymmetries due to duration differences may reverse at very short durations. Of course, care would have to be taken to control for the spectral broadening of brief sounds. We may investigate this in future work.

It is possible that there was some effect of duration on the loudness of the sounds. It is known that pure-tone detection thresholds are reduced for longer sounds up to at least about 250 ms (e.g., Algom, Rubin, & Cohen-Raz, 1989). Note, however, that when the distractors were presented diotically, the interference was reduced, even though the diotic presentation would have increased their loudness (e.g., Algom et al., 1989). Furthermore, even if the effect of duration in our experiments *were* mediated by loudness, this would not invalidate our conclusion that the asymmetries observed are due to attentional selection rather than to peripheral masking.

In Experiments 1–5, we have chosen stimuli distinguished by simple features for which we thought there might be detectors and have shown that strong perceptual asymmetries can occur. In Experiment 2, we investigated whether asymmetries were found for stimuli for which we would not expect feature detectors. In Experiment 6, we perform a similar investigation, but using a procedure similar to that of Experiments 3–5.

Experiment 6

Method

Eight participants who reported normal hearing were tested. All had previously participated in Experiment 2, and 7 had participated in Experiment 4. The aim was to substantiate the idea that the presence of feature detectors was the origin of the asymmetries we had observed, by performing an auditory search with stimuli for which we would not expect neural feature detectors. The general structure of the task was similar to that used for Experiments 3–5, and a schematic of the stimuli is shown in Figure 11. In Conditions 1a and 1b, the target comprised three 75-ms tones of the same frequency, with a 130-ms gap between the first two and a 10-ms gap between the second two, giving a total duration of 365 ms. The distractors also comprised three 75-ms tones at the same frequency, but this time, there was a 10-ms gap between the first two and a 130-ms gap between the second two. In Conditions 2a and 2b, the distractors and targets were the other way around. One, two, four, or eight distractors were present in each trial, in addition to a single sound that on half the trials was an additional distractor and on the other half, a target. The start time of the tones was chosen from a uniform random distribution in the range of 0–1,635 ms. They had a logarithmic distribution in the frequency range of 262–4192 Hz. Each tone was presented at 75 dB SPL.

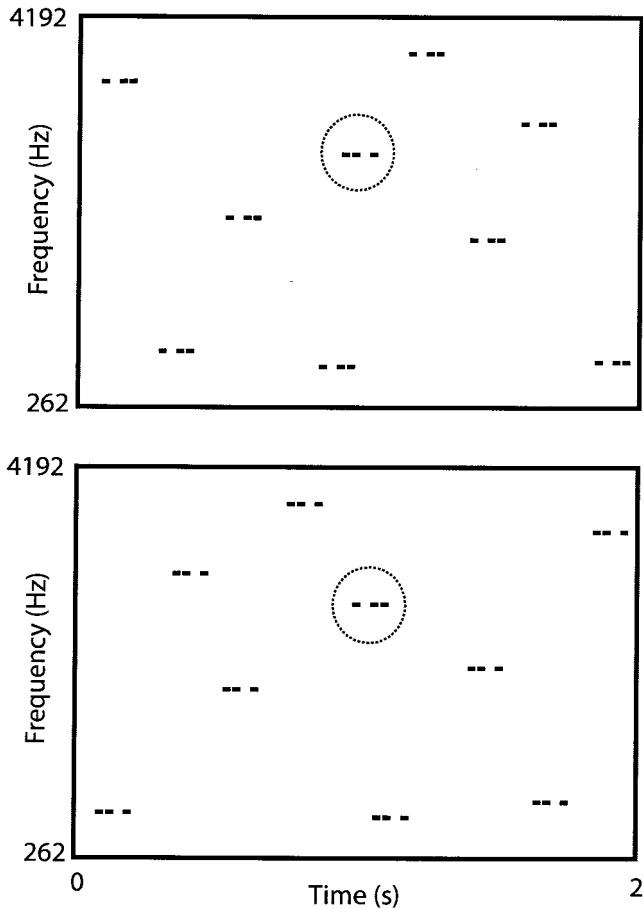


Figure 11. A schematic of example stimuli in Experiment 6 for Conditions 1a and 1b (top) and Conditions 2a and 2b (bottom).

To investigate whether attentional selection or early sensory limitations primarily determined performance, we also manipulated the binaural format of presentation of the sounds in a similar way to Experiment 5. In Conditions 1a and 2a, all of the sounds were presented to the left ear only. In Conditions 1b and 2b, a single sound, which was either a target or a distractor, was presented to the left ear alone. The remaining sounds were presented to both ears. If early sensory limitations determine performance, then as the same sounds were presented to the ear containing the target in Conditions 1a and 1b there would be little difference in performance between these two, and similarly for Conditions 2a and 2b. However, if attentional selection limited performance, the addition of a copy of the bulk of the distractors in the other ear in Conditions 1b and 2b, which would cause them to be perceived in the center of the head, would improve performance. Four blocks of 128 trials were presented. Two adjacent blocks comprised 64 trials of each of Conditions 1a and 1b randomly intermixed, and the other two comprised adjacent blocks of 64 trials each of Conditions 2a and 2b. The order of presentation of Conditions 1 and 2 was counterbalanced across participants using an AB design.

Results

The results are shown in Figure 12. A repeated measures ANOVA with Huynh-Feldt sphericity correction showed a main effect of binaural format (monotic vs. diotic distractors), $F(1, 7) = 50.7, p < .0001$; a main effect of the number of distractors, $F(1,$

$7) = 5.10, p < .001$; and a binaural Format \times Number of Distractors interaction, $F(3, 21) = 8.93, p < .01$. There was no main effect of target type (long-short vs. short-long gap), $F(1, 7) = 1.09, ns$; Target Type \times Binaural Format interaction, $F(1, 7) = 0.00, ns$; Target Type \times Distractor interaction, $F(3, 21) = 0.743, ns$; or a three-way interaction of binaural format, number of distractors, and target type, $F(3, 21) = 1.312, ns$. Analysis of the criterion c showed a single significant effect of the binaural format, $F(1, 7) = 13.1, p < .01$, with participants more likely to say “no target” in the monaural conditions.

Discussion

No evidence of a perceptual asymmetry was found. Our experiment conforms to the criteria suggested by Frick (1995) for situations in which it is valid to accept the null hypothesis. A null hypothesis, that there is no asymmetry for such stimuli, is plausible. We made a “good effort to find an effect”: The experiment was designed using a paradigm for which positive results were found in Experiments 3–5. Care was taken to avoid floor or ceiling effects by choosing conditions that elicited a range of performance. Despite these precautions, no evidence of an asymmetry was found. That these stimuli, which were designed to be unlikely to have specific neural feature detectors, did not elicit an asymmetry further supports our hypothesis that it is the presence of feature detectors for stimuli that leads to perceptual asymmetries in audition.

General Discussion

Perceptual Asymmetries and Attention

Experiments 1–5 revealed a strong perceptual asymmetry, with the salience of target sounds greater when they were frequency modulated (and distractors were not) or when they were longer in duration than distractors, than vice versa. The results support an account of auditory perception, in which sounds are first coded in terms of features such as these and then selected on the basis of analysis either within single-feature representations (easier) or

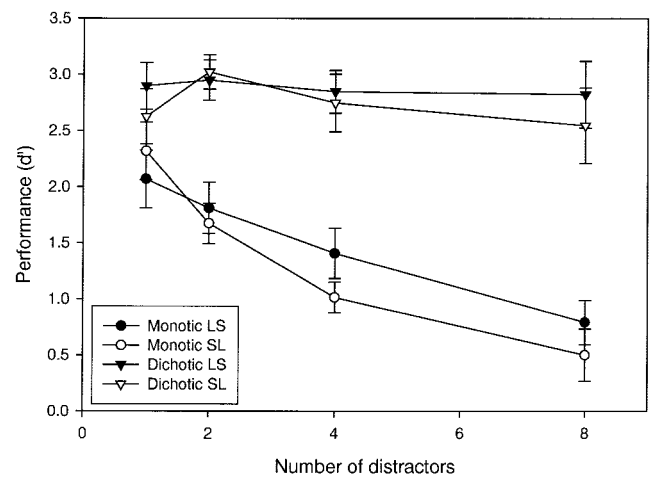


Figure 12. Mean performance (± 1 SEM) of 8 participants in Experiment 6. LS = long-short gap; SL = short-long gap.

across different representations (more difficult). That sounds are coded in terms of features is consistent with predictions made by cognitive analyses (Anderson et al., 1977; Howard & Ballas, 1981; Tou, 1981); neurophysiological evidence (Rauschecker, 1998; Rees & Møller, 1983; see also discussion of Experiment 5); and electrical scalp recordings (Deacon et al., 1998). That discrimination and selective attention then acts on these feature maps would be expected from an analogy with visual models. This leads to the general prediction that when an auditory target has an extra feature that the distractors do not, or a greater magnitude of activation on some dimension, it will be detected more easily, even when explanations based on peripheral masking can be ruled out.

In the introduction, we referred to grouping and selection of sounds as all-or-nothing processes. It might be better to use instead a slightly different description often used in vision based on *competition* between items for further processing (e.g., Bundesen, 1990; Cave & Wolfe, 1990; Duncan & Humphreys, 1989). Sounds that match the target would be more likely to be selected, and processing of distracting sounds would be more likely to be inhibited. Sounds that are perceptually grouped (perhaps to a greater or lesser extent) tend to be selected or inhibited as a group. Adopting this model leads to the interesting prediction that grouping of nontargets would enhance our ability to reject them as a whole. A model that has been successful in quantitatively predicting the results of many studies of visual attention is Bundesen's (1990) theory of visual attention. A promising line of research might be to adapt this model to the auditory domain.

We should consider the relationship between feature extraction and selection. In vision and in audition, it is generally assumed that feature extraction is a parallel, automatic process. In vision, some authors have argued that feature binding—the process of forming objects by combining information across feature representations—requires selective attention (e.g., feature integration theory; Treisman & Gormican, 1988—see introduction). In audition, however, there is more controversy on whether objects can be formed without attention. A number of authors using electroencephalograph (EEG) with the MMN paradigm (Sussman, Ritter, & Vaughan, 1998, 1999; for review see Näätänen & Winkler, 1999) have argued that auditory objects can be formed without attention. However, other work using behavioral measures (Carlyon, Cusack, Foxton, & Robertson, 2001) has suggested that attention *is* necessary for the formation of auditory objects. Further evidence of the importance of attention even in early stages of auditory processing comes from studies using EEG and magnetoencephalography (Hillyard, Teder-Salejarvi, & Munte, 1998; McDonald, Teder-Salejarvi, & Hillyard, 2000; Sussman, Winkler, Huottilainen, Ritter, & Näätänen, 2002; Woldorff et al., 1993). The findings of the current study that performance was worse when information needed to be combined across feature maps support the idea that this object formation is not an automatic process and that limited-capacity resources deployed by attention are indeed required. These results mirror those found in vision. So, why then is there a discrepancy with the findings from the MMN literature? One possibility is the design of those studies. To manipulate attention, we asked participants simply to ignore the tones and read a book. Hence, participants may have been allocating some portion of attention to the tones, particularly as the distracting task was a purely visual one. In a study that did use a competing auditory task (Carlyon et al., 2001), an effect of diverting attention on object

formation was then found. Another possibility is that the effect of attention is primarily to cause a sequence to be heard as a single stream and that some segregation persists even when attention is subsequently diverted elsewhere. Finally, note that Sussman et al. (2002) have recently shown that attention can modify the MMN response, suggesting that, contrary to previous belief, it does not entirely reflect invariant preattentive processes.

It may also be fruitful to consider the degree of innateness of feature extraction mechanisms. Bregman (1990) has suggested a dichotomy between primitive auditory processes, which he defines as being automatic and innate, and higher order selective processes, which he argues are likely to be learned and consciously applied. Feature analysis might be something that is automatic and perhaps innate, but the selection of the task-relevant target through analysis of the feature maps a consciously applied, learned process. However, the distinction between Bregman's categories is now less than clear. Learning may well play a part in the development of feature detectors. For example, the organization of the auditory cortex is heavily dependent on input during development and can even organize to detect visual features (orientation) if rewired to receive visual input during early development (Sharma, Angelucci, & Sur, 2000).

It is likely that automatic processes will influence selection in a bottom-up manner. There might be exogenous orienting to some stimuli. For example, in competition for further processing, frequency-modulated tones might be weighted more heavily than pure tones, corresponding to a deployment of attention on them. It may be that the selection task becomes easier when automatic exogenous orienting to particular stimuli is helpful for the task in hand and more difficult when automatic orienting is unhelpful and must be overruled by endogenous attention.

Some of our findings can also be framed as a form of release from informational masking. Neff and Callaghan (1988), Watson, Kelly, and Wroton (1976), and Leek, Brown, and Dorman (1991) have described situations in which tasks are impeded by distractors even though peripheral masking (loss of information by interference in the early stages of auditory processing) would not be expected. A number of factors are known to modulate informational masking. Neff and Callaghan (1988) showed that the predictability of the distractors had a strong effect on the degree of masking. Oh and Lufti (1998) developed a model showing that for some patterns of stimuli, masking can actually reduce as the number of distractors increases beyond a certain number, as they form a more homogeneous background with less variation from trial to trial. The stimuli in Experiments 3–6 might be expected to generate substantial informational masking, as the distractors (and target) were highly variable from trial to trial. It appears that if the target contains a unique feature, or greater activation on a particular dimension, attentional processes can aid its selection and reduce informational masking. The informational masking framework does not appear to be appropriate in Experiments 1 and 2, however.

Analogies Between Vision and Audition

In our experiments, we have made an analogy between two spatial dimensions in vision, and frequency and time in audition, rather than distributing the sounds in space. There is good evidence that space does not play a primary role in auditory perceptual

grouping. Differences in spatial location can easily be overridden by similarities in frequency in both simultaneous (Darwin & Hukin, 1997) and sequential grouping (Deutsch, 1975). We can separate sounds even with no spatial information (Cherry, 1953). There is also some evidence that in the absence of perceptual grouping, space does not play such an important role in auditory selective attention, both from studies in normal hearing participants (Gockel, Carlyon, & Micheyl, 1999) and in those with attentional deficits following stroke (Cusack, Carlyon, & Robertson, 2000). There are good reasons why we might expect space to play a less important role in audition than in vision. Even in a controlled environment, the precision of localization in the auditory domain is very much worse (1° – 10°) than in vision. Furthermore, in naturalistic environments, sound is reflected and diffracted by objects (M. Campbell & Greated, 1987), which will further impair auditory localization. In summary, spatial information does not seem to play a vital role in auditory perceptual grouping. However, as a caveat, note that once auditory streams are formed and a spatial location is attributed to them, we can use spatial information to selectively attend to them (Darwin & Hukin, 1999; Driver, 1996).

It remains an open question along which dimensions auditory feature detectors are distributed. Primarily on the basis of studies using EEG recording, Näätänen and Winkler (1999) have argued that auditory feature maps are distributed over time. They proposed “that the integration of the individual sound features is guided by the temporal information extracted from the stimulus event” (p. 827). Certainly, throughout the auditory system, neurons are tuned in time, with a temporal integration of the order of milliseconds at early stages and a hundred milliseconds at the level of auditory cortex. It seems likely that feature detectors will also be tuned in time, but might they also be tuned on any other dimensions? Given the strong tonotopic organization of most auditory processing centers in the brain stem and in primary auditory cortex (e.g., Hackett, Preuss, & Kaas, 2001; Talavage, Ledden, Benson, Rosen, & Melcher, 2000), it seems likely that feature detectors might also be tuned by frequency. They might also be tuned in space, although for reasons discussed in the last paragraph, we feel this is less likely. Extensions of the current paradigm, comparing the effect of distributing stimuli over various dimensions, might be an interesting way of investigating this.

Other Evidence for Auditory Perceptual Asymmetries

The results of Experiments 3 and 4 are closely related to a well-established phenomenon, which also implies that perceptual asymmetries are found in audition. It is known that the detection of FM on a target sound is impaired by the presence of another simultaneous tone *if and only if it is also frequency modulated* (Carlyon, 1992; Moore, Glasberg, Gaunt, & Child, 1991; Wilson, Hall, & Grose, 1990). This effect is known as *frequency-modulation-detection interference* (FMDI) and has been well characterized. Carlyon (2000) pointed out that FMDI reflects a perceptual asymmetry, because the discrimination of the order of presentation of a modulated and a steady tone is disrupted by a modulated distractor but not by a steady distractor. More recently, Gockel and Carlyon (2000) have shown that FMDI occurs not only when the distractors and targets are presented simultaneously but also when they are presented sequentially. We have extended this

finding by showing that interference occurs in suprathreshold tasks (Experiments 3–4), and even when the distractors are separated from the targets by as much as 500 ms (Experiments 1–2).

Although modulation detection interference has usually been explained in terms of fairly low level processes, such as the *modulation filterbank* (e.g., Houtgast, 1989; Dau, Kollmeier, & Kohlrausch, 1997), we propose that the effect is more general in scope and that it is better explained in terms of higher level attentional constructs like those that have been used in vision, such as the efficiency of selective attention.¹ Experiment 5 suggests that we see similar effects when duration is manipulated, and the most parsimonious explanation for Experiments 1–5 is that the interference is at the higher level. This is supported by other evidence in the literature showing that higher level organization of sounds can affect the degree of amplitude-modulation-detection interference, such as the reduction observed when masker and target are gated differently (Hall & Grose, 1991) or when they fall in different perceptual streams (Oxenham & Dau, 2001). However, the existence of higher level interference does not preclude the possibility that there may well also be interference at an earlier stage of processing, such as in the modulation feature detectors.

One possible way in which the link between FMDI and our effects could be further elucidated would be to see whether our effect responds in a similar way to FMDI to the variation of modulation rate. It has been demonstrated in a somewhat different task that different modulation frequencies interfere to different extents (Houtgast, 1989), and it has been concluded that the detectors are tuned, albeit rather broadly, by rate. If a similar pattern of interference was identified in our sequential task, it would suggest that a common mechanism underlies both modulation-detection interference and the perceptual asymmetries we observed for modulated sounds.

Work recently presented in summary form by Asemi, Sugita, and Suzuki (2000) used a task in which sounds were presented in free field from one of seven different loudspeakers. A target sound and between one and six distractors were presented on each trial, and participants were instructed to indicate as quickly as possible the presence or absence of a specified target. It was found that reaction times were quicker when the targets were narrowband noises and the distractors were pure tones than vice versa. They were also quicker for the detection of FM or amplitude modulation tones from pure tones than vice versa. Using a different task and measure, their results are in strong agreement with our own.

The results presented here have some quite general implications for the design and interpretation of perceptual experiments. For example, it is commonly assumed (Hanna, 1992) that forced-choice experiments with a small number of intervals tap primarily sensory limitations. In contrast, the results of Experiments 1 and 2 show that perceptual asymmetries, which we assume mediate performance via attentional mechanisms, can lead to differences in d' as large as 1.5–2. Hence, when designing experiments, we

¹ The concept of the modulation filterbank is usually invoked to account for the existence of amplitude-modulation-detection interference (AMDI). However, because FM interferes with the processing of AM and vice versa (Moore et al., 1991), it is probable that the two effects are mediated by the same mechanism.

should take care to control for perceptual asymmetries as they can affect performance substantially.

In conclusion, the experiments presented here provide strong evidence for the existence of perceptual asymmetries in audition. The results suggest that, in analogy with vision and in concordance with cognitive and neurophysiological evidence, the auditory system codes sounds in terms of their features, hence modulating performance in selective-attention tasks. The paradigms described here may be applied to the search for other auditory asymmetries, candidates for which include the coding of loudness, bandwidth, and even higher level categories such as the voicing of speech sounds.

References

- Algom, D., Rubín, A., & Cohen-Raz, L. (1989). Binaural and temporal integration of the loudness of tones and noise. *Perception & Psychophysics*, *46*, 155–166.
- Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.
- Asemi, N., Sugita, Y., & Suzuki, Y. (2000). Auditory search asymmetry with pure tone, narrow-band noise, amplitude-modulated tone, and frequency modulated tone. *Journal of the Acoustical Society of America*, *107*, 2850.
- Beck, J. (1974). Relation between similarity grouping and peripheral discriminability. *Journal of Experimental Psychology*, *102*, 1145–1147.
- Bieser, A. (1998). Processing of twitter-call fundamental frequencies in insula and auditory cortex of squirrel monkeys. *Experimental Brain Research*, *122*, 139–148.
- Brand, A., Urban, A., & Grothe, B. (2000). Duration tuning in the mouse auditory midbrain. *Journal of Neurophysiology*, *84*, 1790–1799.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- Bundesden, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.
- Campbell, F. W., & Maffei, L. (1970). Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *Journal of Physiology*, *207*, 635–652.
- Campbell, M., & Greated, C. (1987). *The musician's guide to acoustics*. London: J. M. Dent.
- Carlyon, R. P. (1992). The psychophysics of concurrent sound segregation. *Philosophical Transactions of the Royal Society of London*, *336*(Series B), 347–355.
- Carlyon, R. P. (2000). Detecting coherent and incoherent frequency modulation. *Hearing Research*, *140*, 173–188.
- Carlyon, R. P., Cusack, R., Foxtton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 115–127.
- Casseday, J. H., Ehrlich, D., & Covey, E. (1994). Neural tuning for sound duration: Role of inhibitory mechanisms in the inferior colliculus. *Science*, *264*, 847–850.
- Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*, 225–271.
- Chen, G. D. (1998). Effects of stimulus duration on responses of neurons in the chinchilla inferior colliculus. *Hearing Research*, *122*, 142–150.
- Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.
- Cowey, A. (1985). Aspects of cortical organisation related to selective attention and selective impairments of visual perception: A tutorial review. In M. Posner & O. Marin (Eds.), *Attention and performance XI* (pp. 41–62). Hillsdale, NJ: Erlbaum.
- Cusack, R., & Carlyon, R. P. (2000). Auditory pop-out: Perceptual asymmetries in sequences of sounds. *British Journal of Audiology*, *34*(2), 112.
- Cusack, R., Carlyon, R. P., & Robertson, I. H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience*, *12*, 1056–1065.
- Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). London: Academic Press.
- 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 and Performance*, *25*, 617–629.
- Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers. *Journal of the Acoustical Society of America*, *102*, 2892–2905.
- Deacon D., Nousak, J. M., Pilotti, M., Ritter, W., & Yang C. M. (1998). Automatic change detection: Does the auditory system use representations of individual stimulus features or gestalts? *Psychophysiology*, *35*, 413–419.
- Deutsch, D. (1975). Musical illusions. *Scientific American*, *233*, 92–98, 103–104.
- Driver, J. (1996). Enhancement of selective listening by illusory mislocalisation of speech sounds due to lip reading. *Nature*, *381*, 66–68.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.
- Ehret, G. (1997). The auditory cortex. *Journal of Comparative Physiology*, *181*, 547–557.
- Frick, R. W. (1995). Accepting the null hypothesis. *Memory & Cognition*, *23*, 132–138.
- Fuzessery, Z. M., & Hall, J. C. (1999). Sound duration selectivity in the pallid bat inferior colliculus. *Journal of Neurophysiology*, *72*, 1061–1079.
- Galazyuk, A. V., & Feng, A. S. (1997). Encoding of sound duration by neurons in the auditory cortex of the little brown bat, *Myotis lucifugus*. *Journal of Comparative Physiology*, *180*(Series A), 301–311.
- Gardner, R. B., & Wilson, J. P. (1979). Evidence for direction-specific channels in the processing of frequency modulation. *Journal of the Acoustical Society of America*, *66*, 704–709.
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*, 103–138.
- Gockel, H., & Carlyon, R. P. (2000). Frequency modulation detection interference produced by asynchronous and nonsimultaneous interferers. *Journal of the Acoustical Society of America*, *108*, 2329–2336.
- Gockel, H., Carlyon, R. P., & Micheyl, C. (1999). Context dependence of fundamental-frequency discrimination: Lateralized temporal fringes. *Journal of the Acoustical Society of America*, *106*, 3553–3563.
- Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *Journal of Comparative Neurology*, *441*, 197–222.
- Hall, J. W., III, & Grose, J. H. (1991). Some effects of auditory grouping factors on modulation detection interference (MDI). *Journal of the Acoustical Society of America*, *90*, 3028–3035.
- Hanna, T. E. (1992). Discrimination and identification of modulation rate using a noise carrier. *Journal of the Acoustical Society of America*, *91*, 2122–2128.
- He, J., Hashikawa, T., Ojima, H., & Kinouchi, Y. (1997). Temporal integration and duration tuning in the dorsal zone of cat auditory cortex. *Journal of Neuroscience*, *17*, 2615–2625.

- Hillyard, S. A., Teder-Salejarvi, W. A., & Munte, T. F. (1998). Temporal dynamics of early perceptual processing. *Current Opinions in Neurobiology*, 8, 202–210.
- Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *Journal of the Acoustical Society of America*, 85, 1676–1680.
- Howard, J. H., & Ballas, J. A. (1981). Feature selection in auditory perception. In D. J. Getty & J. H. Howard Jr. (Eds.), *Auditory and visual pattern recognition* (pp. 181–196). Hillsdale, NJ: Erlbaum.
- Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55–98). Hillsdale, NJ: Erlbaum.
- Kubovy, M., & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97–126.
- Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics*, 50, 205–214.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467–477.
- Ma, X., & Suga, N. (2001). Corticofugal modulation of duration-tuned neurons in the midbrain auditory nucleus in bats. *Proceedings of the National Academy of Sciences, USA*, 98, 14060–14065.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
- McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, 407, 906–908.
- Moore, B. C. J., Glasberg, B. R., Gaunt, T., & Child, T. (1991). Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 40A, 327–347.
- Muller, C. M., & Leppelsack, H. J. (1985). Feature extraction and tonotopic organization in the avian auditory forebrain. *Experimental Brain Research*, 59, 587–599.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125, 826–859.
- Neff, D. L., & Callaghan, B. P. (1988). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *Journal of the Acoustical Society of America*, 83, 1833–1838.
- Oh, E. L., & Lufti, R. A. (1998). Nonmonotonicity of informational masking. *Journal of the Acoustical Society of America*, 104, 3489–3499.
- Oxenham, A. J., & Dau, T. (2001). Modulation detection interference: Effects of concurrent and sequential streaming. *Journal of the Acoustical Society of America*, 110, 402–408.
- Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, 38, 359–365.
- Pinheiro, A. D., Wu, M., & Jen, P. H. (1991). Encoding repetition rate and duration in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. *Journal of Comparative Physiology*, 169A, 69–85.
- Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology Neurootology*, 3, 86–103.
- Rees, A., & Møller, A. R. (1983). Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hearing Research*, 10, 301–330.
- Rose, G. J., & Capranica, R. R. (1983). Temporal selectivity in the central auditory system of the leopard frog. *Science*, 219, 1087–1089.
- Sharma, J., Angelucci, A., & Sur, M. (1990). Induction of visual orientation modules in auditory cortex. *Nature*, 404, 841–847.
- Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, 789, 130–138.
- Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, 36, 22–34.
- Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, 13, 393–405.
- Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, 150, 225–244.
- Tansley, B. W., & Suffield, J. B. (1983). Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. *Journal of the Acoustical Society of America*, 74, 765–775.
- Tou, J. T. (1981). Feature-extraction approach to auditory pattern recognition. In D. J. Getty & J. H. Howard Jr. (Eds.), *Auditory and visual pattern recognition* (pp. 129–142). Hillsdale, NJ: Erlbaum.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Wakefield, G. H., & Viemeister, N. F. (1984). Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels? *Journal of the Acoustical Society of America*, 75, 1588–1592.
- Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns: II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America*, 60, 1176–1186.
- Williams, D., & Julesz, B. (1992). Perceptual asymmetry in texture perception. *Proceedings of the National Academy of Sciences, USA*, 89, 6531–6534.
- Wilson, A. S., Hall, J. W., & Grose, J. H. (1990). Detection of frequency modulation (FM) in the presence of a second FM tone. *Journal of the Acoustical Society of America*, 88, 1333–1338.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences, USA*, 90, 8722–8726.
- Xia, Y.-F., Qi, Z.-H., & Shen, J.-X. (2000). Neural representation of sound duration in the inferior colliculus of the mouse. *Acta Otolaryngologica*, 120, 638–643.
- Zeki, S. M. (1978). Functional specialization in the visual cortex of the rhesus monkey. *Nature*, 274, 423–428.
- Zeki, S. M. (1981). The mapping of visual functions in the cerebral cortex. In R. Norgren & M. Sato (Eds.), *Brain mechanisms of sensation* (pp. 105–128). New York: Wiley.
- Zhou, X., & Jen, P. H. (2001). The effect of sound intensity on duration-tuning characteristics of bat inferior collicular neurons. *Journal of Comparative Physiology*, 187A, 63–73.

Received February 28, 2001

Revision received October 15, 2002

Accepted January 6, 2003 ■