The Intraparietal Sulcus and Perceptual Organization

Rhodri Cusack

Abstract

■ The structuring of the sensory scene (perceptual organization) profoundly affects what we perceive, and is of increasing clinical interest. In both vision and audition, many cues have been identified that influence perceptual organization, but only a little is known about its neural basis. Previous studies have suggested that auditory cortex may play a role in auditory perceptual organization (also called auditory stream segregation). However, these studies were limited in that they just examined auditory cortex and that the stimuli they used to generate different organizations had different physical characteristics, which per se may have led to the differences in neural response. In the current study, functional magnetic resonance imaging was used to test for an effect of perceptual organization across the whole brain. To avoid confounding physical changes to the stimuli with differences in perceptual organization, we exploited an ambiguous auditory figure that is sometimes perceived as a single auditory stream and sometimes as two streams. We found that regions in the intraparietal sulcus (IPS) showed greater activity when 2 streams were perceived rather than 1. The specific involvement of this region in perceptual organization is exciting, as there is a growing literature that suggests a role for the IPS in binding in vision, touch, and cross-modally. This evidence is discussed, and a general role proposed for regions of the IPS in structuring sensory input. ■

INTRODUCTION

Often, the sound arriving from the environment is a mixture from many sources, but we are only interested in a subset of these. The auditory system initially decomposes the sound in many ways, such as by frequency in the ear and by other features later in processing. The resulting vast array of incoming sensory information is then structured, so that parts originating from different sources are allocated to different streams (Figure 1). This process is often referred to as *auditory stream segregation* or *perceptual organization*, the latter term also being used for the analogous process in vision. Auditory stream segregation exploits commonalities across sounds in the environment. For example, sequential sounds that are similar in their frequency or the pitch they evoke are more likely to have come from the same source than sounds with very dissimilar frequency or pitch. Sound components that begin at the same time or change in amplitude in the same way are likely to come from a common source. These and many other heuristics have been found to be important (see Darwin & Carlyon, 1995; Bregman, 1990, for reviews).

Perceptual organization has a strong effect on how we hear the world (Cusack & Carlyon, 2004). Once streams are formed, we can selectively attend to just one at a time, but can generally switch between them at will.[1]

Listeners find it hard to make comparisons in the temporal relationship or order of sounds across different streams (Bregman & Campbell, 1971). Streaming has a strong effect on rhythm (e.g., van Noorden, 1975), timbre, and pitch (Bregman, 1990). For example, if two people's footsteps are interleaved and we hear them as a single stream, we will hear an average tempo that is twice as fast as if we hear them as two separate streams. How we hear music is strongly affected by streaming, and composers explicitly manipulate cues for perceptual organization to achieve particular rhythms, harmonies, and melodic separation (Huron, 1991).

Clinically, abnormal perceptual organization has been implicated in syndromes such as autism and dyslexia. Frith (1989) influentially proposed that a large component of autism could be described by a cognitive processing style of weak central coherence. An important element of this is a greater tendency to perceive the parts of the sensory scene in detail rather than more integrated percepts. This has been demonstrated for many different types of stimuli, both in vision and audition (e.g., Foxton et al., 2003; Happe, 1996, 1999). This suggests a bias in the default form of perceptual organization in autism. It has been thought for some time that abnormal basic auditory processing may also play a role in dyslexia (e.g., Tallal, Miller, & Fitch, 1993; Tallal, 1980). Recently, Sutter, Petkov, Baynes, and O'Connor (2000) and Helenius, Uutela, and Hari (1999) presented simple tone sequences and found the perceptual organization of dyslexic adults differs from that for normal

MRC Cognition and Brain Sciences Unit
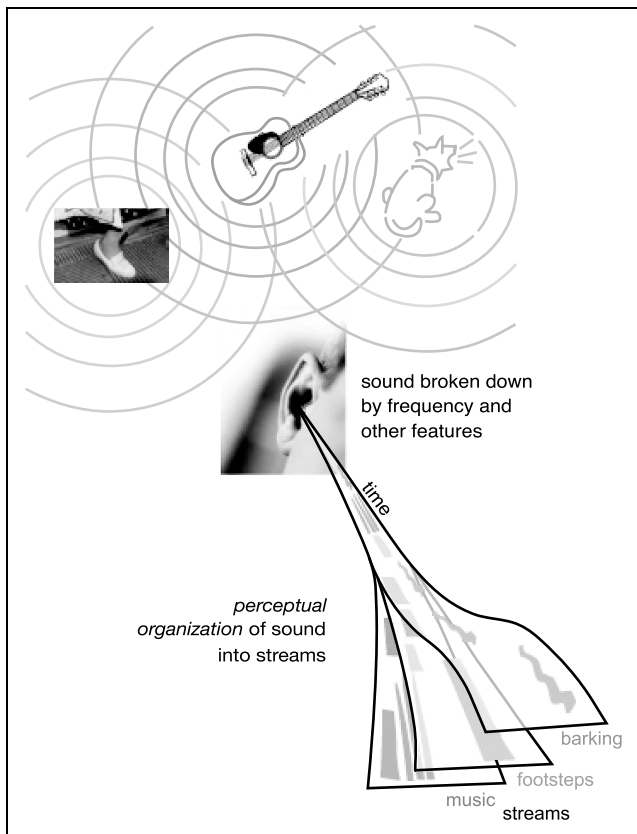
Figure 1. Mixtures of sounds arriving at the ear are first decomposed, and the resulting vast array then structured into a few streams.

controls. This emerging clinical importance places new impetus on the drive to understand the processes that structure our perceptual world.

Although substantial progress has been made in understanding the acoustic characteristics that influence perceptual organization, only a little is known about the neural structures responsible. Two methodologies have been used to investigate the stream segregation of simple sequences of tones, and provided evidence that early cortical auditory areas are involved. First, Fishman, Reser, Arezzo, and Steinschneider (2001) recorded from primary auditory cortex (A1) in macaques while presenting simple tone sequences similar to those often used to investigate human auditory perceptual organization. These were sequences of tones that alternated in frequency (ABABAB...). These can either be heard as a single stream comprising alternating tones or as two streams each comprising just one frequency at half of the tempo (i.e., A–A–A–...and B–B–B–). When the sequence fragments into two streams, only one can be attended to at any one time. In humans, behavioral experiments have established that increasing the rate of presentation, or the frequency separation between the tones, increases the probability that listeners hear the sequence fragment into two streams rather than one. In their study with macaques, for each penetration,

Fishman et al. first identified the sound frequency that evoked a maximal response from the recording site. Alternating sequences were then presented, with the A tones set to this best frequency. As would be expected, a strong response was evoked by the A tones, and a weaker response to the B tones. Fishman et al. studied the response as a function of sequence rate and the frequency separation between the A and B tones. They found that for higher rates, or a larger frequency separations, the response to the B tones was suppressed, relative either to the A tones in the alternating sequence or to B tones when presented alone. Fishman et al. suggested that this suppression is a marker of stream segregation, as it responds in a qualitatively similar way to changes in rate and frequency as in human behavioral studies. They proposed a mechanism for the neural suppression and suggested it may happen in A1 and sketched a model similar to those proposed to explain the human data (McCabe & Denham, 1997; Beauvois & Meddis, 1996), in which streaming is mediated by interactions between cells distributed along a frequency-ordered (tonotopic) representation.

A second line of research suggesting a role for auditory cortical areas in stream segregation comes from the recording of event-related potentials. Sussman, Ritter, and Vaughan (1999) used the mismatch negativity (MMN) paradigm, in which a repeating sequence of sounds is played, interspersed with occasional deviants that are different in some way. It has been found that for many types of stimulus variations, deviance in the sequence triggers neural processes that lead to a pronounced difference in response (the MMN) from the response to the standard stimuli. The MMN has been shown to originate from around auditory cortex in humans, monkeys, cats, and guinea pigs (Sabri, Kareken, Dzemidzic, Lowe, & Melara, 2004; Picton, Alain, Otten, Ritter, & Achim, 2000; Tervaniemi et al., 2000; summaries by Ulanovsky, Las, & Nelken, 2003; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001), although it has also been suggested that there may be an inferior parietal in addition to a temporal lobe component (Levänen, Ahonen, Hari, McEvoy, & Sams, 1996). Sussman et al. presented sequences of tones that alternated in frequency. Occasionally, to form a deviant, one of the tones would be shifted a little higher or lower in frequency. They found that when the sequences were presented slowly, at rates that would not normally cause stream segregation, there was no MMN response to the deviants. However, when they were presented at a faster rate, which would usually lead to stream segregation, an MMN response was evoked. They argued that at the faster rate, as in the absence of deviants, each stream comprises a single frequency, the small frequency shift of the deviants was more noticeable. This indicates that the generator of the MMN, auditory cortex, is affected by the state of stream segregation.

That there are changes in the neural response in auditory cortex that parallel stream segregation is exciting, but it does not imply that there are not other brain regions involved. It does not even imply that this region is critical for streaming—auditory cortex might be reflecting organization that is determined earlier or later in the processing stream. One aim of the current study, therefore, was to use functional magnetic resonance imaging (fMRI) to study correlates of stream segregation across the whole brain, rather than just auditory cortex.

A limitation of the studies of both Fishman et al. (2001) and Sussman et al. (1999) is that in their attempts to create different perceptual organizations, they used different stimuli, and the neural effects they observed might just be because of these physical differences per se, rather than the state of perceptual organization. This possibility is exacerbated by the fact that in neither study was stream segregation actually measured, and so the match between the neural effects and streaming was a qualitative rather than quantitative one. In the current study, a second key aim was to control for physical differences in the stimuli.

One method of studying sound segregation while partially controlling for physical differences was recently presented by Bey and Zatorre (2003). They used "interleaved melodies" in which the notes from two melodies are presented alternately (i.e., first note from melody 1, then first note from melody 2, then second note from melody 1, etc.). They presented a probe melody, which was presented alone, and an interleaved mixture. Listeners were asked whether the probe was present in the mixture. Previous work (Bey & McAdams, 2003) has show that when the probe is presented before the interleaved mixture, it is easier to perform the task than when it is presented afterward. Bey and Zatorre used fMRI to contrast activation in trials when the probe was presented before the mixture with trials in which it was presented after. They found greater activity during the mixture in the inferior frontal operculum when the probe had been presented before the mixture and segregation could be performed. In this design, the stimuli contrasted are similar (with only the order of melody presentation differing), thus, controlling for many physical differences between the conditions. However, although this method does allow control of the physical stimuli, the processes involved are likely to be somewhat different in character from those studied by Fishman et al. (2001) and Sussman et al. (1999). As noted by Bey and Zattore (2003), their paradigm taps into top–down selection processes (*schema-driven selection* in the terminology of Bregman, 1990), whereas Fishman et al. and Sussman et al. both used simple stimuli that are thought to be organized by bottom–up data-driven processes (primitive stream segregation in Bregman's terminology).

In the current study, the aim was to focus upon these more data-driven processes of primitive stream segregation, but still control for physical differences in the stimuli. To do this, we used a novel neuroimaging paradigm in which an ambiguous auditory figure was exploited. In vision, the power of ambiguous figures such as the Necker cube or vase–face illusion to shed light upon higher level processing has led them to receive substantial recent attention (Leopold, Wilke, Maier, & Logothetis, 2002). Like these, our auditory stimulus was compared across physically identical presentations (equating demands on basic processing) but was designed so that across time the percept switched unpredictably between two states of different perceptual organization. fMRI was used to identify brain regions that showed greater activation during one state than the other. As well as possibly observing effects in auditory cortex, we hypothesized that when two streams are perceived rather than one, there will be some brain regions that show increased activation due to the increased load of maintaining two separately bound representations. In contrast, when the entire auditory input is considered as one, no such organization is necessary.

RESULTS

Behavioral

The ambiguous auditory figure used comprised a sequence of tones shown in Figure 2A. It could either be heard as one stream with a galloping rhythm (e.g., repeated triplets of pitch low–high–low) or as two concurrent streams (a low-pitched regular stream and a high-pitched regular stream at half the tempo). The salient difference in rhythm makes it easy for participants to determine which percept they are hearing. They used one of two buttons to indicate what they heard at each moment during the sequences by pressing at the start of each sequence and then whenever they heard a change.

As in previous studies using this paradigm (e.g., Cusack & Roberts, 1999; van Noorden, 1975), 2 streams were heard a greater proportion of the time with larger Δf (see Figure 2B). Note that the percept reported at any one moment was always clearly either one or two streams, but averaging over trials leads to fractional values. The effect of Δf was tested statistically using a repeated measures analysis of variance (ANOVA) on the mean number for each subject across time and trials. Two factors were entered: type of triplet (LHL– vs. HLH–) and frequency separation (1, 3, 5, or 7 semitones). The effect of frequency separation was significant, $F(3,51) = 37.4$, $p < .0005$, but there was no effect of the type of triplet, $F(1,17) = 0.927$, *ns*.

Also, as in previous studies, two streams were heard more often later in the sequences (see Figure 2C). This was tested using another repeated measures ANOVA to compare the mean number of streams at 2 points in

Figure 2. For the correct parameters, these sequences are ambiguous and can be heard with one of two perceptual organizations with different rhythms: (A) left: a characteristic galloping rhythm (*horse*); right, 2 isochronous streams, like Morse code (*morse*). Colored regions correspond to perceptual streams. Within a single sequence, the percept often changes across time. Two streams are more likely to be heard for greater Δf (B) and later in sequences (C).
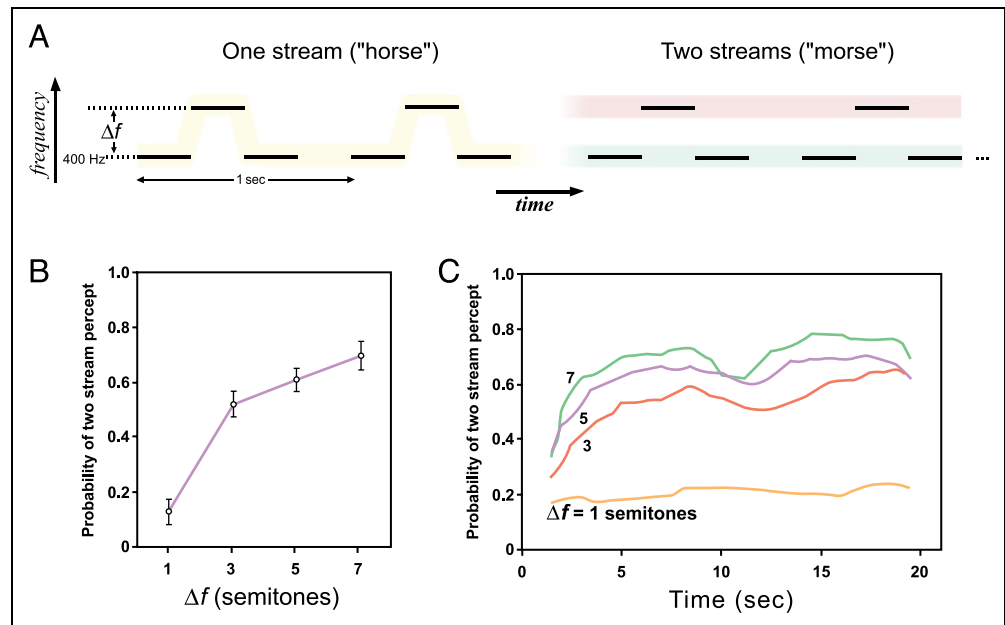
time (early, 2.0 sec; late, 19.5 sec) as a function of frequency separation (1, 3, 5, or 7 semitones). A significant build up over time was found, $F(1,14) = 26.1, p < .0005$. An effect of frequency was also found in this analysis, $F(3,42) = 30.1, p < .0005$, and also a significant Time × Frequency interaction, reflecting the reduced buildup over time for smaller frequency separations, $F(3,42) = 5.0, p < .02$. Note that by visual inspection, there is a dip in the amount of segregation after about 8 sec. This is likely to be because of the distracting sound from the MRI scanner, which performs an acquisition between 7.5 and 10 sec, and may occasionally have distracted the listeners, despite their intention to attend to the tones. Withdrawing attention from a sequence like those used here has been shown to cause its percept to reset to a single stream (Cusack, Deeks, Aikman, & Carlyon, 2004).

Neuroimaging

Presentation of sound and performance of the task led to greater activation than during silence for a network of regions including the superior temporal plane (STP) and anterior intraparietal sulcus (IPS) (Figure 3 and Table 1). In the critical contrast, greater activity was found in the right posterior IPS during the 2-stream percept than the 1-stream percept. To test with greater sensitivity whether activation in anterior IPS and on the STP also varied with the current percept, we also defined spherical ROIs of radius 10 mm around the peaks of activity derived from the sound minus silence contrast. Within the anterior IPS, on the right (and with a trend on the left) greater activity was found when participants reported the 2- than 1-stream percept. However, there was no significant change in activity on the STP. Whole-

brain analysis found no effect of frequency separation or any brain regions that responded more strongly during the 1-stream than 2-stream percept. We also conducted an ROI analysis to investigate with more sensitivity whether there was a decrease in activation in auditory cortex with increasing frequency separation, as
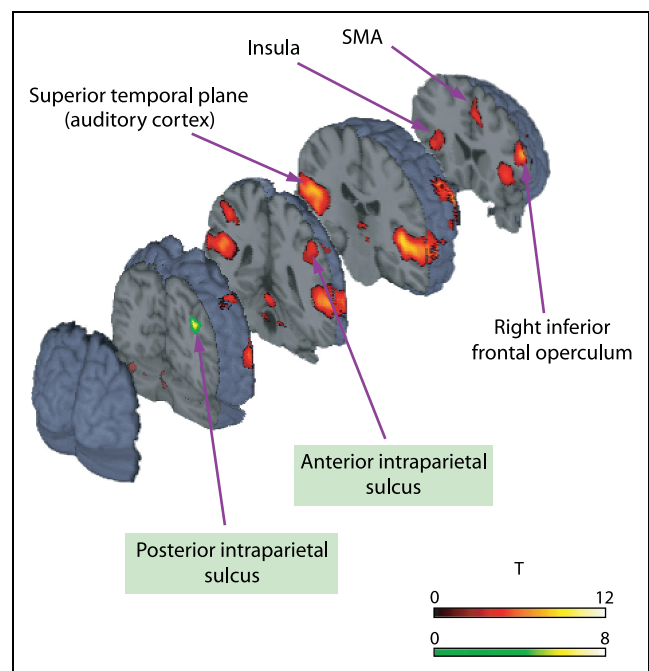
Figure 3. Regions activated when sound was presented and task performed, relative to silence (red–yellow–white colors). When the percept was of 2 streams rather than 1, a right posterior IPS region was activated in the whole-brain analysis (green–yellow–white) and an ROI analysis found activity in the anterior IPS (as indicated by the light green shading).

Table 1. Neuroimaging Results

Region	MNI Coordinates of Primary Peak	Sound Minus Silence		Two Streams Minus One Stream	
		T/t	p (Corrected)	T/t	p (Corrected)
Left superior temporal plane	−48, −28, 12	9.8	<.001	*1.9*	*>.1*
Right superior temporal plane	54, −8, 0	11.3	<.001	*.83*	*>.1*
Right inferior frontal operculum	52, 16, 28	10.1	<.001		
Supplementary motor area	−8, 6, 50	7.2	<.001		
Cerebellum	−34, −60, −34	5.3	<.001		
	4, −62, −16	5.2			
Left insula	−34, 26, 0	6.3	<.001		
Right insula	40, 22, 4	6.0	<.001		
Left anterior intraparietal	−42, −38, 46	6.2	<.001	*2.2 (3.1)*	*.08 (.01)*
Right anterior intraparietal	44, −48, 48	7.5	<.001	*2.9 (3.1)*	*<.02 (<.01)*
Right posterior intraparietal	34, −72, 38			*8.0 (4.6)*	*<.01 (<.0005)*

Italics denote ROI analyses using spheres of 10-mm radii. FDR-corrected SPM$\{T_{17}\}$ were used for whole-brain analyses ($p < .01$), and Bonferroni corrected t values ($p < .05$) for ROIs. Values in parentheses correspond to an alternative statistical model, where rather than removing the effect of frequency separation using a parametric model, 3 separate binary regressors were entered.

might be predicted from the results of Fishman et al. (2001). No such effect was found. In brackets, the results are shown of an alternative model of effects due to frequency, which allowed for nonparametric effects by coding it as 3 binary regressors. This gave similar results, although the left anterior activation did pass statistical significance with this model, but the posterior activation was marginally less significant and did not reach corrected significance in a whole-brain analysis.

DISCUSSION

Three regions in the IPS were found to reflect the state of auditory perceptual organization. In the first subsection, we discuss the relationship between this finding and previous studies of the neural basis of auditory perceptual organization; in the second, its relationship to a number of other studies that have activated the IPS; in the third, some relevant neuropsychology; in the fourth, the limitations of the current study and some ways in which these may be tackled.

Relationship to Previous Studies of Auditory Perceptual Organization

Fishman et al. (2001) and Sussman et al. (1999) argued that primitive stream segregation is mediated by early auditory cortical regions. We did not find any evidence to support this. This difference might be because of the more controlled nature of our stimuli or because of the different recording method used in our experiment.

Our study dissociated the state of perceptual organization from the physical characteristics of the stimuli, whereas in the earlier studies, they were confounded. It might be that if a paradigm similar to ours was used either with neural recording in monkeys, or MMN–EEG (electroencephalogram) in humans, no effects of stream segregation in the absence of physical differences would be observed in auditory cortex. As mentioned in the introduction, neither Fishman et al. nor Sussman et al. took measurements of stream segregation, and so the correlation between their neural findings and stream segregation was qualitative rather than quantitative. As an example, the reduced response to the off-best-frequency B tones when alternated with best-frequency A tones observed by Fishman et al. might reflect some kind of sharpening in representation through lateral inhibition that is not directly related to stream segregation, but which just happens to approximately correlate with it.

Alternatively, there are many differences between the recording methods. Even if perceptual organization does modulate activity in auditory cortex in the absence of physical differences in the stimuli, it might be visible with cell recording or MMN–EEG but not fMRI. Cell recording differs from fMRI in spatial scale and temporal resolution. In Fishman et al.'s (2001) study each contact recorded from a sphere of around 50–100 μm, whereas in fMRI, the resolution (governed by spatial smoothing), is of the order of a centimeter—a difference in volume of around 10^6. For each penetration, Fishman et al. optimized the stimuli to give maximal response, whereas

in fMRI, an average is taken over a few voxels without any optimization. It might be, for example, that although alternating sequences generate less activity at faster rates in the cells recorded from (which were selected to show good tuning to frequency), there were nearby cells that were showing greater activity. In such a case, the net activity change on the scale of an fMRI voxel might be zero. In addition to the different spatial and temporal resolutions, the actual quantity being measured is quite different, and fMRI has been shown to correlate best with the local field potentials rather than the multiunit activity measure as recorded by Fishman et al. (Logothetis, 2002). Similarly, it is difficult to relate EEG and fMRI. For example, even if the differences observed in the MMN are because of perceptual organization (and not frequency differences per se), the measured change in the electrical signal may be a result of changes in synchrony of firing across populations of neurons without any net change in rate. Such differences might not be observable with fMRI. Contrary to this argument, however, is recent evidence from MEG (Singh, Barnes, Hillebrand, Forde, & Williams, 2002) that the synchrony of neural firing may usually be associated with changes in rate. Another limitation that should be borne in mind is that as we did not use cardiac gating to trigger the fMRI acquisition, small structures in the brainstem will have moved around from scan to scan, leading to poor sensitivity.

Given the methodological limitations of fMRI, despite our results, it would be foolhardy to rule out a role for auditory cortex, or other early auditory processing structures, in the perceptual organization of sounds. We have previously argued (Cusack & Carlyon, 2004) for a multilevel model of perceptual organization, in which competition for organization takes place at many stages in auditory processing. In this model, the results from competition at each of these layers would then be integrated to determine the percept. In support of such an idea, acoustic and auditory characteristics best represented at quite different stages in processing have been shown to affect perceptual organization. For example, as already discussed, differences in frequency spectrum (best represented in the auditory system between the cochlea and thalamus) can explain well some aspects of auditory streaming. However, differences in pitch, even in the absence of differences in frequency spectrum, have an effect too (Vliegen & Oxenham, 1999), and although this must be represented separately from frequency, probably later in the processing stream. Differences in location due to interaural time delays (perhaps best represented in the superior olivary complex; Irvine, 1992) also have an effect (Rogers & Bregman, 1998), as do differences in timbre (Roberts, Glasberg, & Moore, 2002; Cusack & Roberts, 2000; perhaps best represented in auditory cortex or later stages; Griffiths & Warren, 2002), and even higher level representations (Scheffers, 1983; probably best represented in other parts of the temporal lobe,

Davis & Johnsrude, 2003). In such a multilevel model, some structure must act to help in the integration of information across these many different regions. We speculate that the IPS plays some role in mediating the combination of information across different representations, and determining the final perceptual organization. However, note that in this model, many other structures are also involved.

Intraparietal Sulcus, Perceptual Organization, and Binding

The IPS has been implicated in several studies investigating the organization of visual information. Its importance in the organization of visual features was demonstrated by Shafritz, Gore, and Marois (2002) using fMRI. They found that regions in the IPS were more activated in feature conjunction tasks than in single feature tasks when multiple objects were presented on the screen at the same time, but not when they were presented sequentially. There are many parallels between perceptual organization and feature binding, both neurally and conceptually (Humphreys, 2001; Müller, Elliott, Herrmann, & Mecklinger, 2001), and their finding could well be related to ours. In our experiment, there was IPS activity when sounds were organized into 2 streams, whereas Shafritz et al. found IPS activity when features presented at different places had to be organized into 2 objects. Donner et al. (2002) also used fMRI, but with a visual search task. They contrasted a conjunction search condition, in which different features have to be bound together, with a feature search condition in which they did not. Although these conditions were matched for difficulty, they found that the IPS was more activated when different features had to be bound together. In a further study using visual search, Wardak, Olivier, and Duhamel (2002) examined the effect of deactivating LIP, a region homologous to the IPS in macaques. They found that when this region was temporarily deactivated through the injection of muscimol, the monkey was still able to perform simple feature searches, but was dramatically impaired in conjunction searches. Further, it was found the deactivation did not affect saccades to single targets, but did affect contralateral saccades when targets were presented on both sides. These findings suggest a role for LIP in the macaque in organizing features in complex scenes.

Interestingly, there is also evidence implicating the IPS in the organization of features in modalities other than vision. Many of these studies have investigated integration across more than one modality presented simultaneously. Calvert (2001) reviews a number of neuroimaging studies and finds that the IPS is often activated during cross-modal integration. This occurs across auditory and visual modalities (e.g., Calvert, Hansen, Iversen, & Brammer, 2001) and across visual and

tactile modalities (e.g., Saito, Okada, Morita, Yonekura, & Sadato, 2003; Macaluso, Frith, & Driver, 2001). There is also evidence that the IPS is involved in integration within the tactile modality (Kitada, Kochiyama, Hashimoto, Naito, & Matsumura, 2003).

The correspondence between this range of literature and our finding suggests that the IPS may be involved generally in the structuring of sensory information across modalities. However, parts of the IPS probably also fulfil other roles, as it comprises several regions with apparently diverse functions such as the suppression of task-irrelevant distractors (Wojciulik & Kanwisher, 1999), mapping salience in the visual scene (Itti & Koch, 2001; Kusunoki, Gottlieb, & Goldberg, 2000), and saccade planning (Colby & Goldberg, 1999). Some of these functions may be partially related to perceptual organization, whereas others may not. Further work will be required to identify how many separable functions exist.

Neuropsychological Evidence Implicating the Parietal Cortex in Object Representation

The IPS is one of the regions damaged in patients with Balint's syndrome, which is characterized by a difficulty in putting the parts of the visual scene together, and simultanagnosia, which is a difficulty in perceiving more than one object. Friedman-Hill, Robertson, and Treisman (1995) found that a patient with Balint's syndrome also showed an abnormally high number of ''illusory conjunctions.'' These are percepts where the features present in the visual scene are incorrectly combined, so that for example when presented with a red ''X'' and a green ''H,'' a green ''X'' is perceived. Bilateral lesions in the intraparietal area led to problems with the correct organization of features, which is consistent with our idea that it is generally involved in structuring sensory information.

Cusack, Carlyon, and Robertson (2000) examined the performance on auditory tasks of patients suffering from unilateral neglect, a deficit that is common following damage to one of a network of regions in the parietal and frontal cortices involved in the control of attention. Although often described as a spatial deficit, there is growing evidence of a number of nonlateralized components to this syndrome (Husain & Rorden, 2003; Robertson, 2001). Cusack et al. (2000) found that patients with unilateral neglect had a specific deficit in making pitch comparisons between with not within sounds. These results are consistent with a model in which, following disruption to parietal function, representation of more than one auditory object becomes difficult. Potentially related deficits have also been identified using visual stimuli. Humphreys (1998) argued that patients with parietal lesions have a deficit in representing multiple objects. He used 2 tasks (counting letters and reading) with words of various lengths, in patients with either temporal or parietal lesions.

Parietal lesions were found to disrupt counting, which he argued required representation of multiple objects, more than reading, whereas temporal lesions led to greater within-object reading deficits. Work with patients with the unilateral neglect has provided two further strands of evidence suggesting deficits in the representation of multiple objects after parietal lesions. This work has used both visual (Manly, Woldt, Watson, & Warburton, 2002; Baylis, Gore, Rodriguez, & Shisler, 2001) and auditory stimulation (Shisler, Gore, & Baylis, 2004).

Limitations and Future Work

There are some possible alternative explanations for our findings. First, we consider the idea that the modulation of activity in the IPS actually reflects the amount of selective attention applied to the stimuli. It has been shown that the application of attention affects auditory stream segregation of sequences like those used here. They are more likely to be perceived as two streams if they have been attended to for some time, compared with the case where they have be presented but not attended to (Cusack et al., 2004; Carlyon, Cusack, Foxton, & Robertson, 2001). Could the change in activity in the IPS actually reflect the application of selective attention, which then leads to stream segregation? We think not, for two reasons. First, there was no request to the listeners to modulate the amount of attention they applied to the sequence and the behavioral results were clean, suggesting the task was performed well throughout. Even when full attention is applied to these sequences, the percept fluctuates randomly, which would have diluted any effect of attention. However, from this alone we cannot rule out this explanation. Second, we have conducted a follow-up study to the current one (Cusack, 2004). In this, stimuli of brief duration (1.2 sec) were presented that were designed to give a percept of either one or two objects perceived simultaneously. In one block, visual stimuli were presented, and in another, auditory stimuli. Unlike the current study, the stimuli with different perceptual organizations were not physically identical, but as a result, the percept was unambiguous. In the visual condition (and with a trend in the auditory), regions in the IPS were activated to a greater extent when two objects were presented rather than one. In this experiment, the modulation of perceptual organization by attention is not a potential explanation as this was determined by the stimulus manipulation.

A second alternative explanation also relates the IPS activation to attention, but rather than it being an ''upstream'' cause of perceptual organization, it could be a ''downstream'' result. It might be that when the percept fragments, listeners can shift attention between the two streams and that activation in the IPS is a result of this shifting. In vision, the IPS has been implicated in shifting attention (Beauchamp, Petit, Ellmore, Ingeholm,

& Haxby, 2001; Nobre, Gitelman, Dias, & Mesulam, 2000; Corbetta, 1998). However, these authors have all shown that covert shifts in visual attention recruit the same mechanisms in the IPS as involved in generating saccadic eye movements, and it is not at all clear how this would relate to shifts in auditory attention between streams. Given this literature, it cannot be ruled out that the IPS activation reflects a mechanism for shifting auditory attention between streams when multiple streams are present. Further experimentation needs to be conducted to investigate this.

Separate auditory processing streams for processing identity ("what") and location ("where") information have been proposed by several authors (Alain, Arnott, Hevenor, Graham, & Grady, 2001; Cusack et al., 2000; Pinek, Duhamel, Cave, & Brouchon, 1989). For example, Alain et al. (2001) used fMRI and EEG to contrast brain activity whereas listeners were performing location and pitch judgments. Although they found greater superior parietal activation for location judgments, it should be noted that they did find substantial bilateral parietal activation for both types of processing. Further work will need to be carried out to relate these findings to ours.

It is not clear why we only observed significant activation in the posterior IPS on the right side. This might just be because of statistical fluctuation, or it might be that the processing really is lateralized to this side. Zatorre and Belin (2001) proposed that a right lateralized system is involved in processing spectral features (as opposed to a left lateralized system for temporal features). However, given that our stimuli involve grouping over time and frequency, it is not entirely clear what would be expected. Again, further work could help relate these findings.

Summary

The BOLD response in regions in the IPS is modulated with the perceptual organization of auditory sequences, even in the absence of differences in stimulation. This result, in combination with a substantial body of literature, suggests that intraparietal regions may play a general role in structuring sensory information. Further experiments could investigate the possible relationships among the organization of information in the auditory, visual, and tactile domains and help us understand selective attention to objects across modalities.

METHODS

Auditory Stimuli and Procedure

The sequences of sounds presented comprised repeating triplets of tones and are commonly used in studies of auditory perceptual organization (e.g., Cusack & Roberts, 2000; van Noorden, 1975). Half of the sequences had the pattern LHL–LHL–... and the other half

HLH–HLH–..., where L represents a low tone, H a high tone, and – a silent interval. The tones and silent intervals were each 125 msec in duration. Two blocks were recorded in each of 18 listeners (11 women, 7 men; mean age 25.9 years with a standard deviation of 6.7 years). The sequences were heard with 1 of 2 perceptual organizations as shown in Figure 2A. To explain the task, a diagram similar to this was shown. When heard as a single stream, a "galloping" rhythm is perceived, which in the instructions to subjects we labeled *horse*. When the sequence splits into two streams, each is isochronous, and the galloping rhythm is lost. In the instructions, this percept was labeled *morse*, as it sounds a little like Morse code. Listeners were asked to report their percept at each moment, by pressing one of 2 buttons. At the start of a sequence, as soon as they knew which percept they heard, they pressed the appropriate button. They were then asked to choose again each time they heard a change in percept. The button mapping was counterbalanced across subjects. As there are some individual differences in propensity to hear one or two streams for different stimuli, and to ensure some balance in response distribution in all subjects, we presented a range of frequency separations. However, at the analysis stage, we partialled out the effect of this. The frequency separation between the low and high tones was 1, 3, 5, or 7 semitones. The mean of the frequency of the two tones (expressed in semitones) was chosen from 0, 3, 6, 9, 12, 15, 18, or 21 semitones above 309 Hz. Over the 2 blocks, 1 sequence of each mean frequency and separation was presented (8 mean frequencies × 4 frequency separations = 32 sequences). The sounds had 25-msec linear rise and fall ramps. They were presented in random order, uniquely generated for each subject. The sequences were 20 sec in duration and separated by 10-sec gaps. They were presented using high quality MR compatible electrostatic headphones (MRC Institute of Hearing Research, Nottingham, UK: www.ihr.mrc.ac.uk/research/technical/soundsystem/index.php). Each block took 16.5 min. The stimuli were presented diotically at 95 dB SPL. To attenuate the sound from the scanner, and reduce the stimulus to a comfortable listening level, listeners wore earplugs (EAR Ultra-tech, attenuation 14.5, 16.7, 18.7 dB at 250, 500, and 1000 Hz).

Neuroimaging: Acquisition

To reduce the interfering effect of noise from the scanner, which can both acoustically mask the sounds and lead to a saturation of auditory cortical areas (Hall et al., 1999), we used a sparse imaging fMRI design. In this, the onsets of the acquisition were separated by 10 sec. The scan took 2.5 sec, and so there were 7.5 sec periods of silence between them in which the acoustic stimuli could be presented without masking. The predicted hemodynamic response due to the scanner noise

Figure 4. The fMRI design. (A) The volume acquisition time is shown by vertical bars (one per slice). The dotted line shows the hemodynamic response to the scanner noise, as calculated using the canonical hemodynamic response function from SPM. Note that in this sparse imaging design, the response to the sound of the previous scan will have mostly died away by the time of the next one. (B and C) The solid lines show the "boxcar" time courses for the percept throughout the presentation of an example trial. In (B), a value of 1 indicates the 1-stream percept and a value of 0 indicates the 2 streams, and vice versa in (C). The dotted lines show the convolution of these boxcars with the hemodynamic response, which were the regressors in the fMRI model. In this example, it can be seen that scan X mostly reflects the 1-stream percept and scan Y mostly reflects the 2-stream percept.
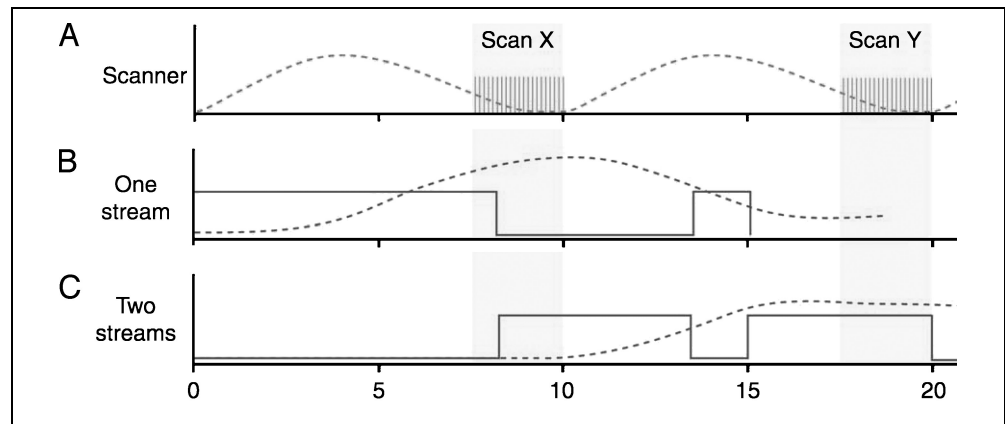
is shown in Figure 4A. For functional imaging, we used an echo-planar imaging (EPI) sequence with 21 slices, a matrix size of 128 × 128, TA = 2.5 sec, and TE = 27.5 msec. In each block, 99 scans were acquired. In each subject, we also acquired maps of the magnetic field using two 3-D SPGR acquisitions with the same TR but slightly different echo times (TE = 7 and 16.104 msec). After phase unwrapping (Cusack & Papadakis, 2002), the phase difference at each voxel between these 2 acquisitions is proportional to the deviation in the z component of the B0 magnetic field strength because of the presence of the head. This information was used to correct distortions in the echo-planar functional images during analysis.

Neuroimaging: Analysis

The initial dummy scan was discarded. Analysis was performed using SPM99 (www.fil.ion.ucl.ac.uk/spm). This began with motion correction. Then, distortions in the EPIs were corrected using a custom toolbox as described in Cusack, Brett, and Osswald (2003) (see also www.mrc-cbu.cam.ac.uk/Imaging/Common/fm_background.shtml). Next, the data were normalized (each subject's images warped into a common space using a nonlinear fitting procedure) and then spatially smoothed with a Gaussian kernel (full width half maximum 8 mm).

A standard SPM regression model was used to partition components of the blood oxygen level-dependent (BOLD) response at each voxel. Two columns coded the perceptual organization heard at each moment (1 binary column set to 1 when percept was *horse* and 0 otherwise; another binary column set to 1 when percept was *morse* and 0 otherwise). At the start of each sequence, it was assumed that the percept before the first response was the same as the first response. Models from an exemplar trial are shown in Figure 4B and C. Another column coded for button presses (an appropriately

timed event). Another column coded frequency separation, Δf (parameter with value corresponding to $\Delta f -$ mean$(\Delta f) = \Delta f - 4 = -3, -1, 1,$ and 3 for spacings of $\Delta f = 1, 3, 5,$ and 7 semitones, respectively). This parametric model allowed us to test for brain regions that showed decreased response with increasing frequency separation, as might be predicted in auditory cortex from the findings of Fishman et al. (2001). Finally, a column coded the through each sound sequence (first or second scan coded with -1 or $+1$, silent interval with 0). The first 4 columns were convolved with the hemodynamic response, but the parametric columns were not. The MarsBar toolbox (marsbar.sourceforge.net/) was used to perform region-of-interest (ROI) analyses. Random effects analyses were used throughout. An alternative model was performed to check whether some nonlinear effect of frequency separation might affect the 2-stream minus 1-stream contrasts. In this alternative model, rather than a parametric column for frequency, 3 separate binary columns were entered that partialled out any differences due to this factor, even those that are nonlinear. The results from this alternative analysis, which was otherwise identical, are shown in parentheses in Table 1.

Note

1. Note, however, that attention can be drawn in a bottom–up manner, such as to highly salient events in the environment.

These are referred to as *exogenous* (as opposed to *endogenous*) shifts in attention.

REFERENCES

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12301–12306.

Baylis, G. C., Gore, C. L., Rodriguez, D., & Shisler, R. J. (2001). Visual extinction and awareness: The importance of binding dorsal and ventral pathways. *Visual Cognition, 8,* 359–380.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., & Haxby, J. V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. *Neuroimage, 14,* 310–321.

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America, 99,* 2270–2280.

Bey, C., & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 267–279.

Bey, C. E., & Zatorre, R. J. (2003). Recognition of interleaved melodies. An fMRI study. *Annals of the New York Academy of Sciences, 999,* 152–154.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge: MIT press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11,* 1110–1123.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14,* 427–438.

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 and Performance, 27,* 115–127.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22,* 319–349.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, U.S.A., 95,* 831–838.

Cusack, R. (2004). Structuring our world: fMRI shows primary cortices and the intraparietal sulcus are involved in perceptual organization. *Journal of Cognitive Neuroscience, Supplement: CNS 2004 Programme, 223.*

Cusack, R., Brett, M., & Osswald, K. (2003). An evaluation of the use of magnetic field maps to undistort echo-planar images. *Neuroimage, 18,* 127–142.

Cusack, R., & Carlyon, R. P. (2004). Auditory stream segregation inside and outside the laboratory. In J. G. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 16–48). New York: Academic Press.

Cusack, R., Carlyon, R. P., & Robertson, I. H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience, 12,* 1056–1065.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 643–656.

Cusack, R., & Papadakis, N. (2002). New robust 3-D phase unwrapping algorithms: Application to magnetic field mapping and undistorting echoplanar images. *Neuroimage, 16,* 754–764.

Cusack, R., & Roberts, B. (1999). Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception, 28,* 1281–1289.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics, 62,* 1112–1120.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. Moore (Ed.), *Hearing* (pp. 387–424). San Diego, CA: Academic Press.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience, 23,* 3423–3431.

Donner, T. H., Kettermann, A., Diesch, E., Ostendorf, F., Villringer, A., & Brandt, S. A. (2002). Visual feature and conjunction searches of equal difficulty engage only partially overlapping frontoparietal networks. *Neuroimage, 15,* 16–25.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research, 151,* 167–187.

Foxton, J. M., Stewart, M. E., Barnard, L., Rodgers, J., Young, A. H., O'Brien, G., & Griffiths, T. D. (2003). Absence of auditory "global interference" in autism. *Brain, 126,* 2703–2709.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science, 269,* 853–855.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford, UK: Blackwell.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25,* 348–353.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., & Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping, 7,* 213–223.

Happe, F. G. (1996). Studying weak central coherence at low levels: Children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry, 37,* 873–877.

Happe, F. G. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Science, 3,* 216–222.

Helenius, P., Uutela, K., & Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain, 122,* 907–913.

Humphreys, G. W. (1998). Neural representation of objects in space: A dual coding account. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences, 353,* 1341–1351.

Humphreys, G. W. (2001). A multi-stage account of binding in vision: Neuropsychological evidence. *Visual Cognition, 8,* 381–410.

Huron, D. (1991). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception, 9,* 135–154.

Husain, M., & Rorden, C. (2003). Non-spatially lateralized mechanisms in hemispatial neglect. *Nature Reviews Neuroscience, 4,* 26–36.

Irvine, D. R. F. (1992). Physiology of the auditory brainstem. In A. N. Popper & R. R. Fay (Eds.), *The mammalian auditory pathway: Neurophysiology* (pp. 153–231). Springer-Verlag: New York.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Kitada, R., Kochiyama, T., Hashimoto, T., Naito, E., & Matsumura, M. (2003). Moving tactile stimuli of fingers are integrated in the intraparietal and inferior parietal cortices. *NeuroReport, 14,* 719–724.

Kusunoki, M., Gottlieb, J., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: The representation of abrupt onset, stimulus motion, and task relevance. *Vision Research, 40,* 1459–1468.

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature Neuroscience, 5,* 605–609.

Levanen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex, 6,* 288–296.

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences, 357,* 1003–1037.

Macaluso, E., Frith, C. D., & Driver, J. (2001). Multimodal mechanisms of attention related to rates of spatial shifting in vision and touch. *Experimental Brain Research, 137,* 445–454.

Manly, T., Woldt, K., Watson, P., & Warburton, E. (2002). Is motor perseveration in unilateral neglect ''driven'' by the presence of neglected left-sided stimuli? *Neuropsychologia, 40,* 1794–1803.

McCabe, S. L., & M. J., D. (1997). A model of auditory streaming. *Journal of the Acoustical Society of America, 101,* 1611–1621.

Müller, H. J., Elliott, M. A., Herrmann, C. S., & Mecklinger, A. (2001). Neural binding of space and time: Spatial and temporal mechanisms of feature–object binding. *Visual Cognition, 8,* 273–285.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). ''Primitive intelligence'' in the auditory cortex. *Trends in Neurosciences, 24,* 283–288.

Nobre, A. C., Gitelman, D. R., Dias, E. C., & Mesulam, M. M. (2000). Covert visual spatial orienting and saccades: Overlapping neural systems. *Neuroimage, 11,* 210–216.

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-otology, 5,* 111–139.

Pinek, B., Duhamel, J. R., Cave, C., & Brouchon, M. (1989). Audio-spatial deficits in humans: Differential effects associated with left versus right hemisphere parietal damage. *Cortex, 25,* 175–186.

Roberts, B., Glasberg, B. R., & Moore, B. C. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America, 112,* 2074–2085.

Robertson, I. H. (2001). Do we need the ''lateral'' in unilateral neglect? Spatially nonselective attention deficits in unilateral neglect and their implications for rehabilitation. *Neuroimage, 14,* S85–S90.

Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception and Psychophysics, 60,* 1216–1227.

Sabri, M., Kareken, D. A., Dzemidzic, M., Lowe, M. J., & Melara, R. D. (2004). Neural correlates of auditory sensory memory and automatic change detection. *Neuroimage, 21,* 69–74.

Saito, D. N., Okada, T., Morita, Y., Yonekura, Y., & Sadato, N. (2003). Tactile–visual cross-modal shape matching: A functional MRI study. *Brain Research, Cognitive Brain Research, 17,* 14–25.

Scheffers, M. T. M. (1983). *Sifting vowels: Auditory pitch analysis and sound segregation.* The Netherlands: Groningen University.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 10917–10922.

Shisler, R. J., Gore, C. L., & Baylis, G. C. (2004). Auditory extinction: The effect of stimulus similarity and task requirements. *Neuropsychologia, 42,* 836–846.

Singh, K. D., Barnes, G. R., Hillebrand, A., Forde, E. M., & Williams, A. L. (2002). Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *Neuroimage, 16,* 103–114.

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.

Sutter, M. L., Petkov, C., Baynes, K., & O'Connor, K. N. (2000). Auditory scene analysis in dyslexics. *NeuroReport, 11,* 1967–1971.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9,* 182–198.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences, 682,* 27–47.

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping, 10,* 74–79.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience, 6,* 391–398.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* The Netherlands: Eindhoven University of Technology.

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America, 105,* 339–346.

Wardak, C., Olivier, E., & Duhamel, J. R. (2002). Saccadic target selection deficits after lateral intraparietal area inactivation in monkeys. *Journal of Neuroscience, 22,* 9877–9884.

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron, 23,* 747–764.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946–953.