

Representation of Multiple Sound Sources in the Owl's Auditory Space Map

Terry T. Takahashi and Clifford H. Keller

Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403

The barn owl's inferior colliculus contains a retina-like map of space on which a sound generates a focus of activity whose position corresponds to the location of the sound source. When there is more than one source of sound, the sound waves sum and may generate spurious binaural cues that degrade the auditory image. We investigated the signal conditions under which neurons in the owl's auditory space map are able to resolve two simultaneously active sound sources. We recorded from space map neurons responding to sounds from a pair of speakers separated in azimuth by 45° and mounted on a rotatable arm. Stimuli consisted of a sum of sinusoids or pseudorandom noise bursts emitted simultaneously and at equal overall levels. The characteristics of the sounds in each speaker were varied, and the neuron's response was plotted as a function of the speaker pair's position. When the speakers emitted different sets of summed sinusoids, the cells responded to each speaker separately; that is, the cells were able to resolve two separate targets. However, when the speakers emitted identical summed sinusoids generating binaural cues that were identical to those of a single phantom source between the two speakers, the neurons responded when the speakers were on either side of their receptive fields. By manipulating the amplitude at which each speaker emitted the various frequencies, we could control the position, number, and size of the phantom sources detected by the cell. The cells also resolved two separate sources when they emitted noise bursts that were statistically independent or temporally reversed versions of one another. Since the overall spectra of such waveforms are identical, we suggest that the space map relies on differences between noise bursts that exist over brief time spans.

[Key words: binaural, inferior colliculus, interaural, nucleus mesencephalicus laterale pars dorsalis, sound localization, spatial hearing]

The natural acoustical environment is often cluttered with objects that emit sounds at the same time. Yet, we hear sounds with distinct identities, each attributable to a particular source. One of the basic principles underlying this process is the ability

to determine the location of sound sources (Cherry, 1953; Sayers and Cherry, 1957; Bregman, 1990). Conventional neurophysiological studies have made substantial progress in elucidating the mechanisms by which neurons encode the binaural cues generated by a single source in a quiet environment. Less attention, however, has been paid to the more natural condition in which there are multiple acoustical sources.

In the visual system, the external world is projected onto the retina and topographic maps are built from the preservation of retinal topology. In the auditory system, however, spatial attributes must be derived from the comparison of waveforms in the left and right ears. When two sources are simultaneously active, each ear receives the sum of the waveforms from the two sources. If the two sources' sounds are similar, the process of waveform summation generates binaural signals that can lead to serious errors in localization. Given the complexity of the natural environment, one would expect the spectra of various natural sources to overlap at any given time. How, then, does the auditory system manage to localize each source accurately?

The barn owl (*Tyto alba*), a nocturnal bird of prey renowned for its ability to localize sound, is uniquely suited for the neuroethological study of the representation of multiple sound sources (Payne, 1971; Konishi, 1973). The external nucleus of its inferior colliculus (ICx) contains auditory neurons, called space-specific neurons, that are highly selective for the location from which a sound emanates (Knudsen and Konishi, 1978a). As in the visual system, space-specific neurons are arrayed in a topographic fashion so that an acoustical event is represented as a focus of neural activity whose position in ICx corresponds to the position of the source in space (Knudsen and Konishi, 1978a). The ICx can therefore be said to contain the neural images of sound sources, and thus gives us the opportunity to examine how a complex acoustical environment is represented at a neural level.

In the present report, we describe some of the basic conditions under which the owl's auditory space map is able to represent two, simultaneously active sound sources separated along the horizontal axis.

Coding of auditory space

For frequencies that are most relevant to its ability to localize sound, the owl uses interaural time differences (Δt) and interaural level differences (ΔI) to determine the azimuth and elevation of a source, respectively (Knudsen and Konishi, 1979; Moiseff, 1989). These two binaural cues are processed in functionally independent and anatomically nonoverlapping pathways that extend from the cochlear nuclei to the central nucleus of the inferior colliculus (ICc), the sole and direct source of

Received Sept. 28, 1993; revised Dec. 17, 1993; accepted Jan. 20, 1994.

We thank John Martin for help with computer software and Petr Janata for technical help. Petr Janata and Walter Heiligenberg provided many helpful discussions and criticisms. This study was supported by grants from the National Institutes of Health, the Medical Research Foundation of Oregon, and the Office of Naval Research.

Correspondence should be addressed to Terry T. Takahashi and Clifford H. Keller at the above address.

Copyright © 1994 Society for Neuroscience 0270-6474/94/144780-14\$05.00/0

inputs to the space map of ICx (Knudsen, 1983; Moiseff and Konishi, 1983; Sullivan and Konishi, 1984; Takahashi et al., 1984; Wagner et al., 1987; Takahashi and Konishi, 1988a,b; Takahashi et al., 1989). For the present study, in which two sources of sound were separated along the horizontal axis, the most pertinent binaural cue is Δt . In the owl, Δt is computed from the interaural differences in the phase angles of a sound's spectral components. The details of the mechanism by which interaural phase difference ($\Delta\phi$) is computed have been reviewed elsewhere (Sullivan and Konishi, 1986; Konishi et al., 1988; Takahashi et al., 1989; Carr and Konishi, 1990), and it suffices to say here that in ICc, neurons selective for $\Delta\phi$ and frequencies are organized into tonotopic columns according to their preferred $\Delta\phi$ and frequency values. Within such a column, a cell's preferred $\Delta\phi$ changes as a linear function of the cell's best frequency (Wagner et al., 1987). Thus,

$$\Delta\phi = \Delta t \times F, \quad (1)$$

where F is the cell's best frequency. Note that the rate at which the preferred $\Delta\phi$ changes with frequency within a column is Δt . Neurophysiologically, this means that all neurons of a column are activated by a particular Δt . Columns in ICc are systematically arrayed forming a topographic representation of Δt .

Figure 1A diagrams ICc as a block, subdivided into boxes representing neurons. Frequency is mapped from top to bottom, and Δt , from left to right. By convention, positive Δt values mean that the right ear's sound leads. In the figure, the ICc is shown responding to a single source broadcasting a sound containing 5000, 6000, and 7000 Hz, from a location that gives rise to a Δt of +30 μsec . According to Equation 1 (above), the source would activate ICc cells tuned to 5000 Hz and 0.15 cycles, 6000 Hz and 0.18 cycles, and 7000 Hz and 0.21 cycles. These cells are arranged tonotopically within the column that represents +30 μsec and are represented by the darkly shaded boxes in Figure 1A. Note that in Figure 1A there are also two lightly shaded boxes to the left. Phase is a cyclic variable, and a given phase value recurs for Δt values that differ by integer multiples of the stimulus period. Thus, these boxes represent activity in the 6 and 7 kHz lamina of ICc whose preferred $\Delta\phi$ corresponds to a Δt that is one period removed from 30 μsec . Thus, the leftmost lightly shaded box represents activity in the cell with a preferred $\Delta\phi$ that corresponds to -137 μsec (=30-167 μsec ; 167 μsec is the period of 6 kHz) and the other lightly shaded box is at -112 μsec (=30-142 μsec). The analogous activity for the 5 kHz band, which would be at -170 μsec , is not visible.

Cells in a column project convergently onto a cluster of space-specific neurons endowing them with a selectivity for the Δt preserved by the column and a sensitivity to a broad range of frequencies (Wagner et al., 1987). In Figure 1A, a vertical stack of boxes is shown activating a space-specific neuron. Thus, space-specific neurons can be most strongly excited by a multifrequency sound whose components' $\Delta\phi$ values correspond to the Δt conserved within its input column. Note that the lightly shaded boxes of Figure 1A do not fall along a vertical column. Neurophysiological evidence suggests that such "decolumnated" activity will only weakly activate space-specific neurons (Takahashi and Konishi, 1986).

The mechanism described above, by which binaural time delays are derived from binaural phase differences, was first demonstrated in the midbrain and olivary nuclei of mammals (Rose et al., 1966; Geisler et al., 1969; Goldberg and Brown, 1969; Yin and Kuwada, 1983; Yin et al., 1987; Yin and Chan,

1990; see also Yin et al., 1989, for review) and has also been proposed to operate in man (Stern et al., 1988).

Superposition of waveforms and the representation of multiple sources

How does the inferior colliculus represent two simultaneously active sources of sound? A sound wave composed of a single frequency can be represented as a vector, also called a phasor, that rotates in the complex plane (Fig. 2). The phasor's length is the sound's amplitude and its angle with respect to the real axis (abscissa) is its phase. Since frequency is the rate of change of phase with time, it is equivalent to the vector's rate of rotation about the origin. The phasor representation is particularly useful for visualizing the process of waveform summation. Figure 2 uses a "snapshot" of phasors at an instant in time to show how a single frequency from two sources adds in each ear (Bauer, 1961; Blauert, 1983). The paths from each source to each ear are shown (Ll, Lr, Rl, Rr) at the top. The vector diagrams in the left and right columns (Fig. 2A-C) represent the tone from each source (solid lines) and the resultant of phasor addition (dashed lines) in each ear. The drawing assumes that the speakers are at the owl's eye level and that there are no echoes. For simplicity, the two speakers shown are equidistantly placed from the owl's midline, but the same principles apply anywhere in the horizontal plane. The binaural cues, $\Delta\phi$ and Δl , can be computed for this frequency by subtracting, respectively, the angles of the left and right resultants and the lengths of the left and right resultants. Because $\Delta\phi$ and Δl are processed independently, the cues are *not* computed by first performing a vector subtraction on the resultants, and then computing the angle and length of the remainder (Moiseff and Konishi, 1983; Sullivan and Konishi, 1984; Takahashi et al., 1984; Takahashi and Konishi, 1988).

Figure 2A shows that when two speakers emit identical tones at the same amplitude, the resultant phasors in the left and right ears are identical. There is no left-versus-right difference in the resultants' angles or lengths, and thus there are no interaural differences in phase or level. Perceptually and physically, this should be equivalent to a single source located halfway between the two loudspeakers.

If the tones emitted from the two speakers are not of equal loudness, the vector addition in each ear results in an interaural *phase* difference (Fig. 2B). Again, a phantom source appears between the two speakers, but it is now closer to the louder source. In Figure 2B, the right source is shown as being twice as loud, and after the phasors sum, the right ear hears a signal that is phase advanced relative to that in the left ear. Note that despite the *interspeaker* level difference, there is no resultant *interaural* level difference. For a given interspeaker level difference the relationship between the frequency and the advance in $\Delta\phi$ is approximately linear for the range of frequencies used here, causing the phantom's Δt to be biased toward the louder source.

If there is a delay between the two speakers' tones, the phase relationships of the components (solid lines) will differ between the left and right ears. In Figure 2C the right source is shown leading by an amount of time equivalent to one-quarter of a cycle ($\pi/2$). The vector summation produces a longer resultant (dashed lines) in the left ear than in the right, which, by convention, is designated as a negative Δl . The owl should perceive a phantom source located below the horizon, because left-louder signifies locations below the horizon to the owl, and halfway

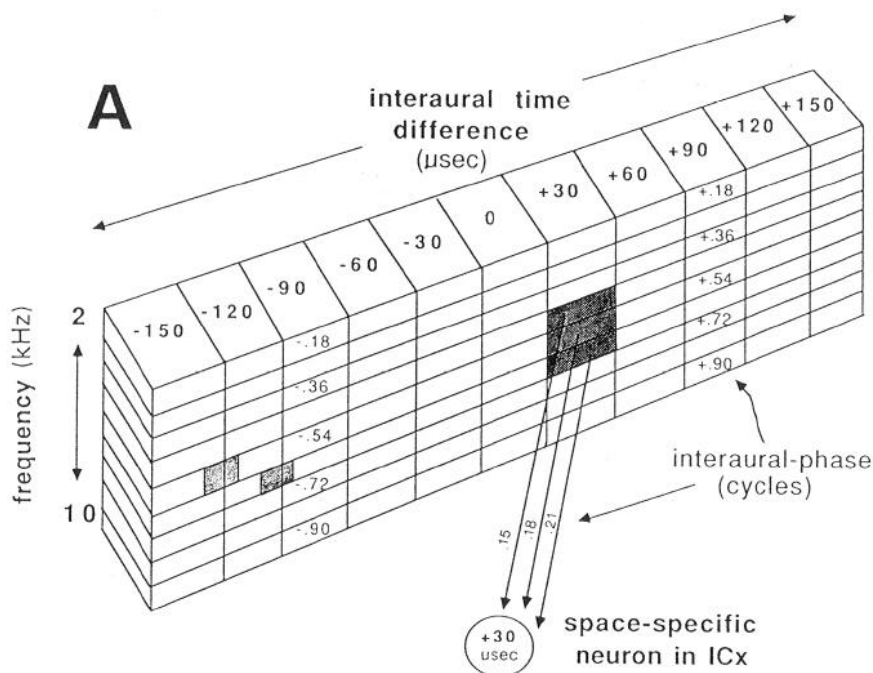
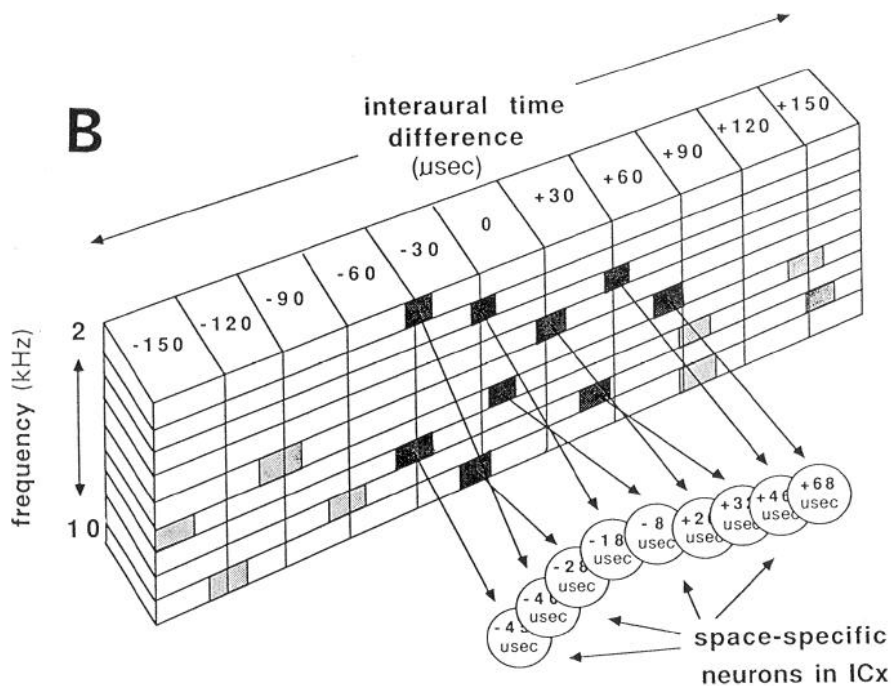


Figure 1. Schematic representation of the barn owl's ICc and its projection to the space-specific neurons of the ICx. Note that ICx is innervated by that portion of ICc that maps contralateral space plus a limited portion of ipsilateral space. Two stimulus configurations are represented, for which areas of high evoked activity are *darkly shaded*. **A**, A single sound source emits a sound consisting of 5, 6, and 7 kHz sinusoids, from a location that gives rise to a Δt of 30 μsec . In ICc, cells tuned to these frequencies and the $\Delta\phi$ equivalents (Eq. 1) of the 30 μsec Δt , are shown activated. In ICx, a space-specific neuron innervated by the column is activated. **B**, Two sources emit nine frequencies in common. The interspeaker level difference for each spectral component differs, however, causing the apparent spatial location for each component to be different. In ICc and ICx, the neural activity is scattered.



between the two speakers, because $\Delta\phi$ is zero (Knudsen and Konishi, 1979; Moiseff, 1989). If the speakers emit a broadband sound, and one sound is slightly delayed, the summation within each frequency band will result in a different Δt for each frequency. These different Δt values may indicate slightly different elevations, because the relationship between Δt and elevation depends on frequency in a manner determined by the shape and size of an owl's head and external ears. Thus, the phantom image could be blurred, primarily, in the vertical plane.

Given the organization of the ICc, we can describe how two

simultaneously active sources might be represented in the space map of ICx. If two speakers each emit spectral components that are not present in the other speaker, we can expect two distinct columns and two sets of space-specific neurons to be activated in ICc and ICx, respectively. If some frequencies are common to both speakers and others are unique to each source, one might expect to see three columns of activation in ICc. The frequencies unique to each source will excite columns that represent the two sources' Δt values, and the common frequencies will mix and excite columns that represent the average of the two sources' Δt

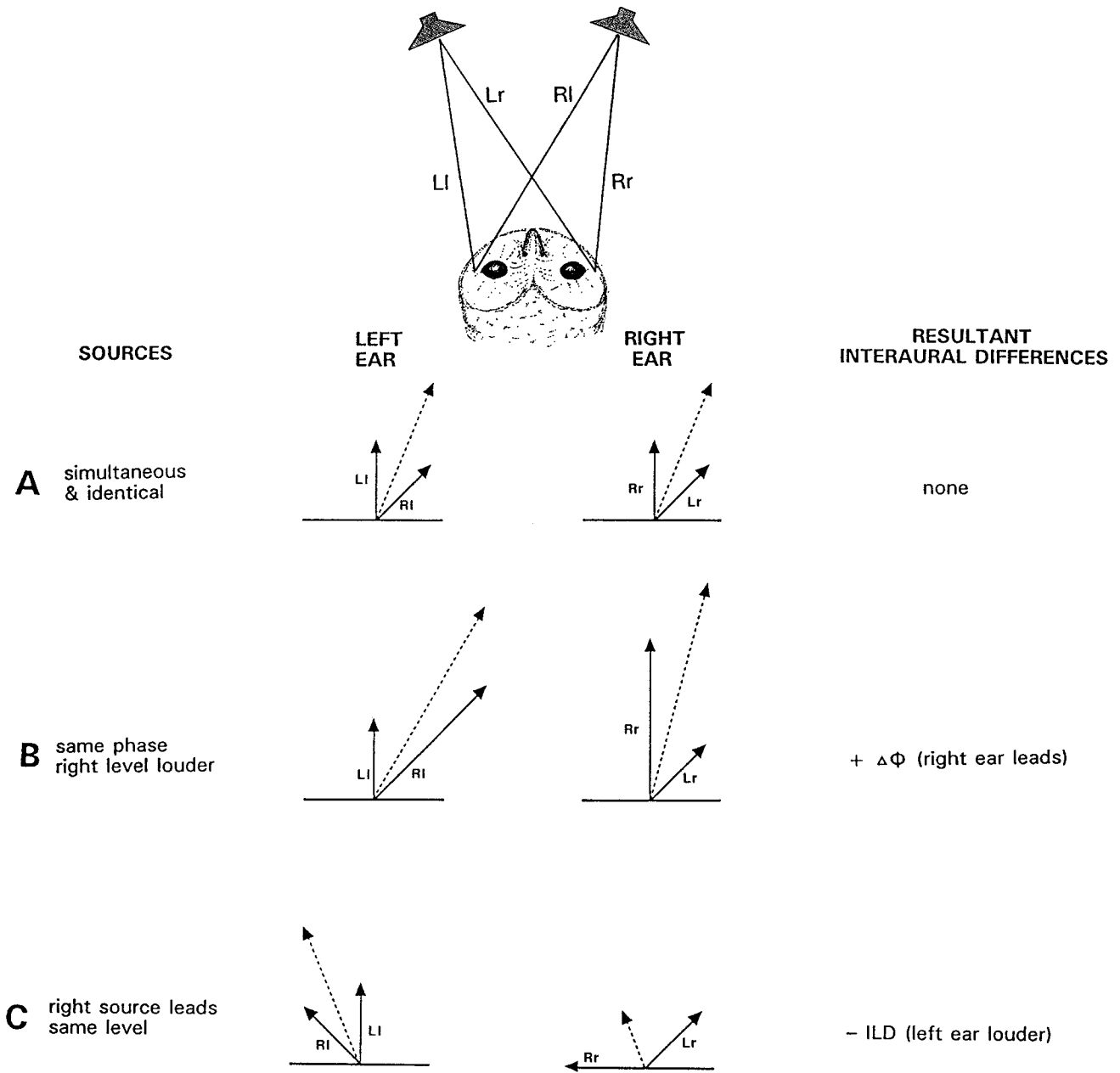


Figure 2. Phasor diagrams depicting the vector summation in the complex plane of two sound sources as received by each ear. The sound sources are equidistant from the owl's head and placed symmetrically about the midline on a horizontal plane. A phasor represents the sound's level by the vector's length and its phase by the angle relative to the positive-going real axis (*x*-axis). Each vector will spin about the origin at a rate equal to the tone's frequency, in this case all frequencies are the same and thus all vectors have the same rate of rotation. Vectors from each source to each ear are plotted as *solid lines* and are denoted as in the *drawing at the top*; for example, *Lr* indicates the sound from the left source to the right ear, and so on. The vector resulting from the summation of these tones is shown as a *dashed line* in each plot, and interaural differences in phase and amplitude can be evaluated by comparing, respectively, the angles and lengths of these *dashed vectors* for each ear. Three stimulus conditions are represented. In *A*, sounds are emitted simultaneously and in phase from each speaker and at the same level. The resultant vectors at the two ears are identical. *B* shows the situation when both tones are emitted simultaneously and in phase, but the tone from the right speaker is twice as loud as from the left. The resultant vectors at each ear are of the same length, but the right vector is phase-advanced. In *C*, the emitted tones' levels are identical, but the right source is phase advanced relative to the left. In this case, the resultants show no phase difference but rather the left ear receives a louder sound.

values. If two sources contribute unequally to a common frequency, the resulting $\Delta\phi$ for that frequency is closer to the $\Delta\phi$ of the louder source, and the activity in ICc will be similarly biased. This is generalized in Figure 1*B* to the condition in which there are multiple frequencies, each with a different interspeaker amplitude difference. The result is a scattered pattern of activity in ICc, which may produce a low level of discharge over a wide

area of the ICx, giving rise to the perception of a spatially diffuse image.

Materials and Methods

Surgery and animal care. All procedures were carried out under a protocol approved by the Institutional Animal Care and Use Committee of the University of Oregon. Data were gathered from seven adult barn

Table 1. Frequency components used for summed sinusoids

Frequency (Hz)	Column A	Column B	Column C
2000	X	X	
2360	X		
2722		X	X
3080			X
3345	X	X	
3706	X		
4052		X	X
4411			X
4772	X	X	
5113	X		
5461		X	X
5819			X
6192	X	X	
6545	X		
6910		X	X
7258			X
7601	X	X	
7966	X		
8333		X	X
8670			X

owls. An owl was anesthetized by intramuscular injection of ketamine (0.1 ml/hr; Vetalar, 100 mg/ml; Parke-Davis) and diazepam (0.05 ml/hr; Diazepam C-IV, 5 mg/ml; LyphoMed) and given a prophylactic dosage of ampicillin (0.2 ml, i.m.; Polyflex, 20 mg/ml; Aveco). The owl was wrapped in a hot water blanket and placed into a stereotaxic device that held the owl's head tilted downward at a 45° angle. After topical application of lidocaine (2% lidocaine HCl; Xylocaine, Astra Pharmaceuticals), the scalp was cut to expose the rostral cranium where a stainless steel plate was cemented. Once the plate was firmly affixed, it served as a means of holding the head in a fixed position allowing removal of the ear bars and bite plate. A hole about 0.7 cm² was opened in the skull with sterile rongeurs, lined with silicone grease, and filled with saline to prevent desiccation.

To minimize stress on the bird, we recorded in two 15–18 hr sessions, separated by a 4–7 d recovery period. After the first session, the craniotomy was closed with dental cement (Vitrabond, 3M Inc.) and the scalp sutured and coated with an antibacterial cream (Bacitracin-Neomycin-Polymyxin Ointment, E. Fougerra and Co.). The bird was then given 0.2 ml of dexamethasone and 0.2 ml of vitamin B complex and placed into a temperature-controlled recovery room.

Birds were killed at the end of the second session with an overdose of pentobarbital (Nembutal, Abbott Laboratories), exsanguinated (0.9% NaCl in 0.1 M phosphate buffer), and fixed (3.7% paraformaldehyde in 0.1 M phosphate buffer, pH 7.4) by transcardial perfusion. The brains were removed and processed for histology to confirm the locations of electrode penetrations.

Stimulus generation and data analysis. All recordings were carried out in a 1.83 × 1.83 × 1.83 m echo-attenuated booth. Sound stimuli were presented through a loudspeaker array mounted on a 90 cm arm that could be rotated about the owl's head in the horizontal plane. The array comprised three loudspeakers (Alpine 6010EX), equally spaced along the chord of a horizontal arc (45°). Search stimuli were presented through the loudspeaker in the center of this array. After isolation and initial characterization of a cell's response, test stimuli were presented through either or both of the lateral speakers.

All stimuli consisted of digitally synthesized noise bursts and summed sinusoids (Table 1) with starting phases of 0 radians. The digital waveforms were converted into an analog form at a rate of 50,000 samples per second (Data Translation 2821; Modular Instruments Inc. M308). Stimuli were trapezoidally enveloped (5 msec on/off time, 100 msec duration), scaled to 20–30 dB above neuronal threshold with programmable attenuators (Modular Instruments Inc.), and amplified stereophonically (McIntosh 754).

Our experiments required the generation of a phantom image by emitting identical sounds from the two lateral speakers simultaneously. To get proper waveform summation it is necessary that the two lateral speakers be equidistant from the center of the bird's interaural axis. To achieve this condition, one peripheral speaker was mounted on a mechanical stage so that its distance to the bird was adjustable. We tested and adjusted for equidistance in two ways. First, we reasoned that if the two sources were equidistant, there should be no frequency-dependent ΔI values. We therefore compared the two ears' waveforms by inserting matched miniature microphones (Knowles Electronics, BT 1954) into the bird's ear canals, emitting broadband noises through the speaker pair, and generating a Lissajous figure on an oscilloscope using the output of the microphones. The spread of the Lissajous figure reflected ΔI values over all of the frequencies and could be minimized by adjusting the distance between the bird and the stage-mounted loudspeakers. This somewhat cumbersome technique was used to validate a simpler method that made use of the neuronal response to various interspeaker delays. The speaker array was positioned so that the two peripheral speakers were on either side of a space-specific neuron's receptive field, as determined previously with a single speaker. In such a configuration, a strong phantom source, generated halfway between these speakers, should stimulate the cell most effectively. We sought to maximize the cell's response by emitting identical broadband noises from each speaker while systematically delaying the sound from one speaker under control of a high-speed clock (Modular Instruments M110). Plots of firing rate as a function of interspeaker delay peaked at the delay required to compensate for the difference in travel time between the bird and each of the two sources. The stage-mounted loudspeaker was then moved toward or away from the bird and the neuron retested to see whether the peak was now at or near zero interspeaker delay. This latter method was employed for all data reported below.

Action potentials from isolated neurons were recorded with fine tungsten electrodes (Frederick Haer Co.; 10 M Ω) and the time of their occurrence relative to stimulus onset was written to computer disk. Spike rates are reported as the difference of the stimulus-evoked rate less the spontaneous rate recorded immediately before the auditory stimulus and are the average over 10 presentations for each stimulus condition. Statistical differences in spike rates between stimulus conditions were tested using the Wilcoxon signed-ranks test.

Results

We recorded extracellularly from individual space-specific neurons in the ICx while producing sounds from one or more speakers on the speaker array. The array was placed at different azimuths, and neuronal firing was plotted as a function of the array's position. An array position of 0° is directly in front of the bird, negative angles are to the bird's left and positive angles to its right. The plot, called a "spatial response profile," gives an idea of the role of a cell in representing the acoustical environment under any given stimulus condition. Thus, when two speakers simultaneously emit a stimulus configuration that allows the neurons to resolve each speaker as separate entities, the cell should respond twice, as the array is positioned at different azimuths and each speaker is sequentially placed in its receptive field. A bimodal spatial response profile would be obtained. Assuming that all space-specific neurons have similar properties, spatial response profiles allow us to infer the pattern of activity across the space map in ICx.

Figure 3 presents spatial response profiles for four cells (A–D) when stimulated with a sound burst consisting of a sum of 10 sinusoids. These cells represent the range of responses of our sample of 30 cells, the least typical of which is shown in Figure 3D. For each cell, we first defined the receptive field by plotting its response to the summed sinusoids emitted from the central speaker (left column). For these plots, speaker position and array position are synonymous. The response for a typical cell is shown in Figure 3A. This cell showed a unimodal response profile, responding vigorously with the array positioned at -10° (10° to the left of midline), and only weakly when it was positioned

elsewhere. These single-speaker profiles, which closely resemble the original receptive-field plots of Knudsen and Konishi (1978a), serve as a reference for comparison of the same cell's response to various stimuli presented through the two speakers at the ends of the speaker array.

As discussed above, binaural cues from frequencies that are unique to only one sound source will be unchanged by those from other sources and are thus expected to reflect the sources' locations accurately. To test the response of space-specific neurons to this stimulus, we presented a sound burst composed of the sum of 10 different frequency sinusoids simultaneously from the two speakers. The central speaker was silent for these tests. The left and right speakers emitted the spectral components indicated in columns A and C of Table 1, respectively. With each sound composed of a sum of 10 different sinusoids, we expected these two sources to be represented as separate peaks on a cell's spatial response profile. The center column of Figure 3 (open circles) shows spatial response profiles for this stimulus configuration for the same four neurons (A–D). The profile shown in Figure 3A shows two distinct peaks of activity corresponding to array positions that placed either the right (25°) or the left (–25°) speaker into the neuron's receptive field. The distance between the peaks is consistent with the separation between the two outer speakers. The response minimum in the center corresponds to the array position in which the two peripheral speakers straddled the receptive field. Similar spatial response profiles were obtained from the other three representative neuron's whose responses are shown in graphs B–D of the same column. Note that the double-peaked profiles are smaller than the peak obtained with a single speaker (shown in left column) from the same neuron. Thus, the contrast between the peaks of the profile and regions off of the peak was reduced for multiple sources ($p < 0.0001$; $n = 29$ cells).

In contrast to the 10 unique frequencies discussed above, identical sinusoids emitted simultaneously from two sources will add and give rise to a phantom image located midway between the two speakers. Spatial response profiles for the same four neurons (Fig. 3A–D), generated with the two outer speakers emitting 10 identical frequencies (Table 1, column A) is shown by the graphs with solid circles in the center column of Figure 3. As exemplified in Figure 3A (center column), a strong response was elicited when the array was positioned such that the two speakers were on either side of, that is, straddled, the receptive field, bringing the phantom source directly into the receptive field. In contrast, when a speaker was centered on the receptive field, the response was greatly diminished. These profiles obtained with 10 common frequencies strongly resembled single-speaker receptive fields. In our sample of 30 neurons, the curves generated with 10 common frequencies were not statistically different ($p > 0.2$) from those obtained with single speaker stimulation in either location or height of the peak. There was a strong tendency, however, for the profiles generated by the 10 common frequency stimulus to be broader as measured by their variance assuming a normal curve ($n = 20$; $p < 0.002$). The phantom source thus appears to act much as a real single loudspeaker. The broader spatial extent of the phantom may reflect slight mismatches in the amplitude response of the two loudspeakers (less than 4 dB in any spectral band), which could cause the phantom image to disperse as explained above.

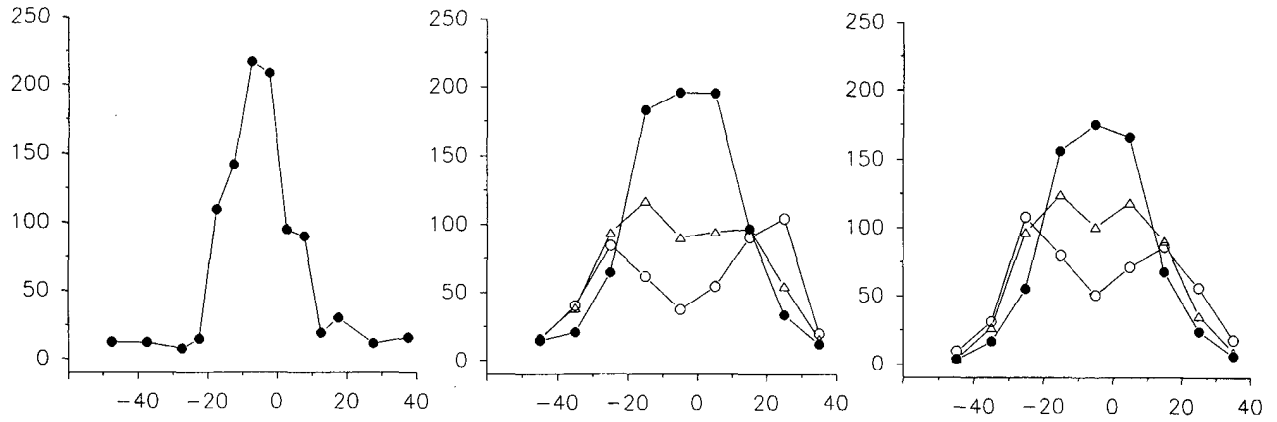
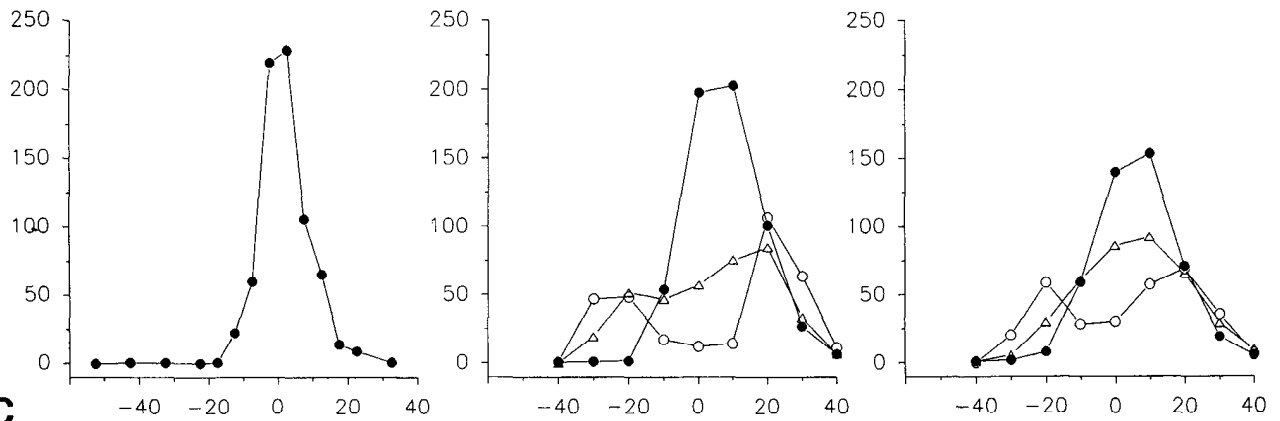
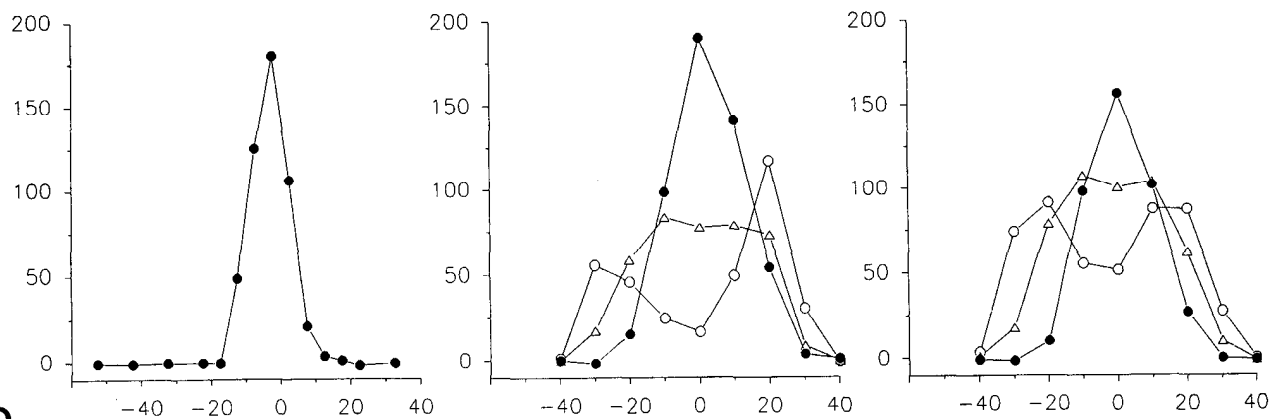
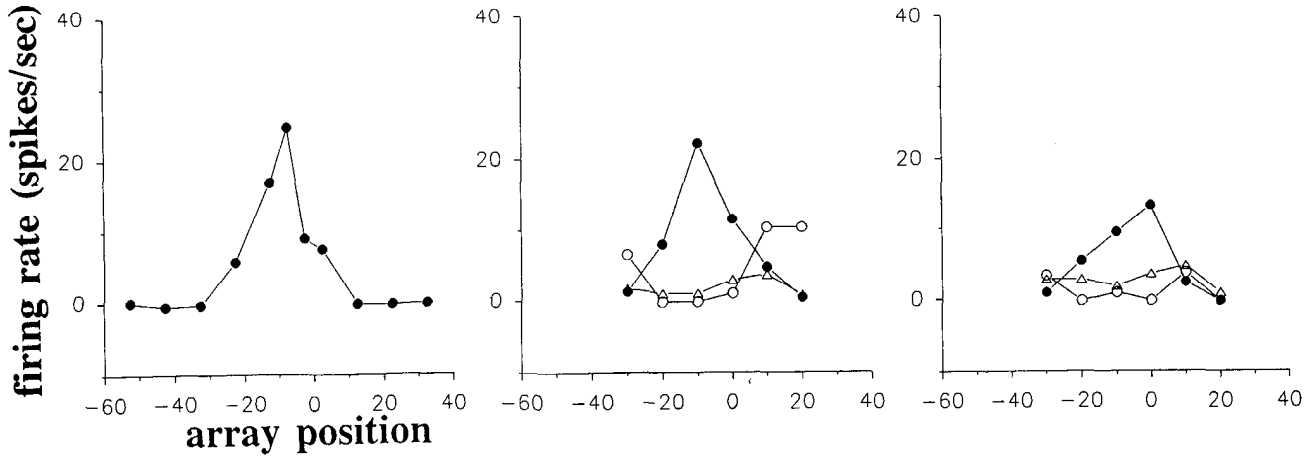
The differences in response to 10 common and 10 unique frequencies described above appear to hold for our sample of 30 cells ($p < 0.0001$). This is illustrated in Figure 4, in which

the firing rate of a neuron obtained with one speaker on the receptive field is plotted against its rate obtained when the two speakers straddled the receptive field. When the two speakers' sounds were identical (solid circles) the cells' firing was stronger with the two speakers straddling the receptive field. When the sounds were different (open circles) the firing was stronger when one of the two speakers was directly on the receptive field.

The third set of profiles in the center column of Figure 3A–D (triangles) represents a case intermediate to the two described above. For these plots, each speaker emitted a sum of 10 sinusoids, five of which were common to the two speakers and five of which were unique to each speaker (Table 1, columns A and B). The common frequencies were expected to produce a phantom source in the middle, whereas the unique frequencies were expected to allow the cells to resolve the actual sources. With this stimulus configuration, however, the spatial response profiles did not show three peaks of activity. Instead, a broadly elevated firing level between the two speaker locations with one or more moderate peaks was seen. When peaks were present, they generally fell between the central peak generated in response to 10 common frequencies (solid circles) and the lateral peaks obtained with 10 unique frequencies (open circles) (Fig. 3A,C, middle column). Although our sampling of the spatial response profile may not have been fine enough, it appears that the neurons did not fully parse the sound into three separate sources.

As explained above, by emitting common frequencies from two speakers and adjusting the amplitude of each frequency component at each speaker, phantom sources for a given frequency can be located anywhere between the two speakers. Applying this method to the spectral components shown in Table 1, column A, we created three phantom stimulus configurations and obtained spatial response profiles with each of them (Fig. 3, right column). In the first configuration, the 10 sinusoids had interspeaker amplitude differences that ranged from ± 2 dB to 6 dB (negative dB values mean that the right speaker was louder), thereby clustering the phantoms midway between the actual speakers ("center"). In the second, the 10 sinusoidal components had a wider range of interspeaker amplitude differences ($\pm 26, 15, 10, 5,$ and 2 dB), and were chosen such as to spread the phantoms evenly between the two speakers ("spread"). In the final configuration, five phantoms were clustered near each of the real speakers ("split") by assigning one of two, roughly equal but opposite interspeaker level differences to each spectral component ($\pm 14, 16, 18, 21,$ and 26 dB). For each stimulus, the spatial response profiles closely resembled the responses expected based on the mixing of common frequencies, weighted by their levels in each source. The right column of Figure 3 shows that the neurons responded strongly to one central location when the phantom sources were clustered in the center (solid circles), but to two peripheral targets when the phantoms were split (open circles). With the spread configuration, the spatial response profiles had a single broad peak, indicating that the neurons could not discriminate between individual sources. As best exemplified in Figure 3C, however, the total width of the profile obtained with the spread configuration (triangles) is narrower than the total width of the profile obtained in the split configuration (open circles) even though the interspeaker amplitude differences for some of the frequencies in the spread configuration would cause the phantoms to extend from speaker to speaker.

The tests described above were all performed using a sound composed of a sum of 10 sinusoids. Each sound's spectrum was

A**B****C****D**

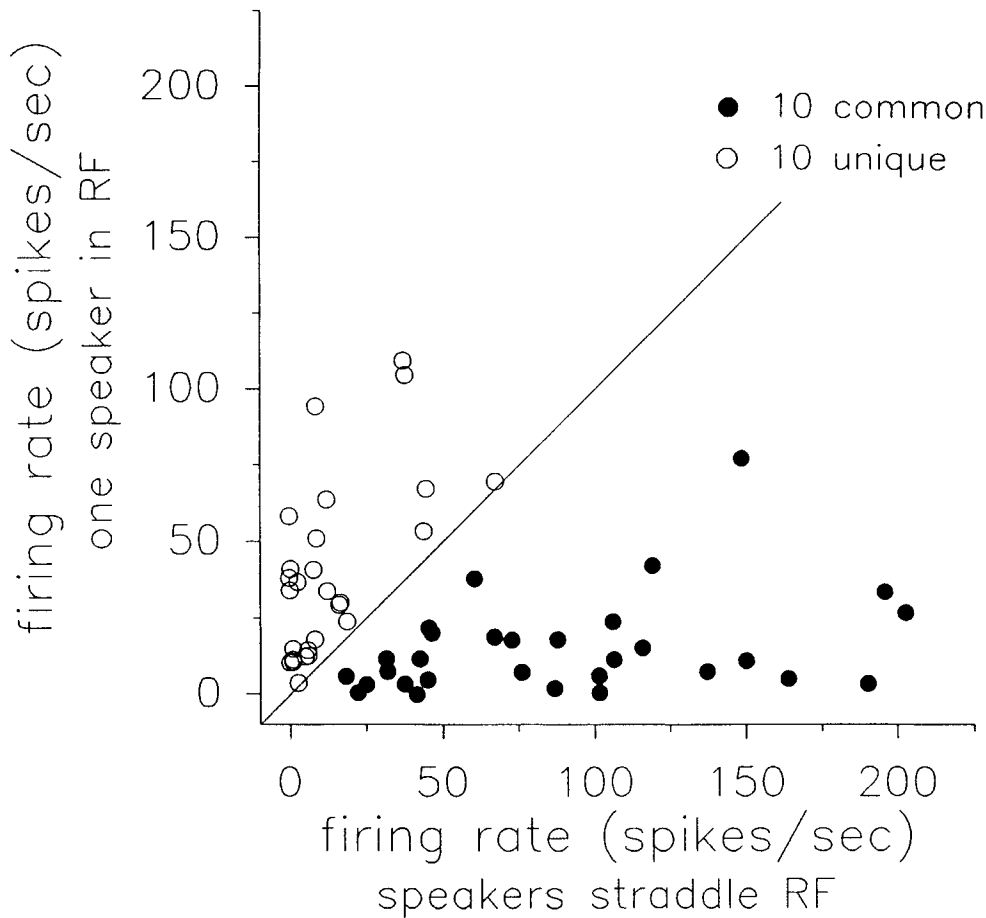


Figure 4. Responses of 30 cells recorded within the ICx and stimulated with sounds from two speakers when one speaker is centered on the cell's receptive field (*ordinate*) or when the two speakers straddle the receptive field (*abscissa*). Each cell was stimulated both with 10 common (*solid circles*) and with 10 unique frequencies (*open circles*).

constant over time (excepting the amplitude effects of ramping) and consisted of a comb-like array of frequencies. Naturally occurring sounds, such as the rustling produced by prey, are generally broadband and quite dynamic; frequencies come and go at random and with randomly varying amplitudes. In the tests that follow we used broadband, random noise bursts, which more closely mimic natural sounds and more thoroughly challenge the capabilities of the auditory space map. The same basic principals of sound mixing apply, but the resultant vectors for each spectral component change randomly from moment to moment.

As with the sum-of-sinusoids tests above, we first constructed spatial response profiles for each neuron with one speaker. Profiles from four cells (different cells from those shown in Fig. 3) are shown in the left column of Figure 5A–D. They closely resembled those obtained with the sum-of-sinusoids (Fig. 3A–D, left column), and displayed single discrete peaks when the active speaker was located in the neuron's receptive field. When the two speakers simultaneously emitted identical noise bursts,

which was expected to cause a phantom source to appear halfway between the two real sources, spatial response profiles displayed a single strong peak of activity when the two speakers straddled the receptive field (Fig. 5A–D, right column, solid circles). There was no apparent response when the individual speakers were placed in the cells' receptive fields. These curves were again strongly similar to those obtained with a single speaker when comparing the height and location of the peaks ($p > 0.1$; $n = 44$). Figure 5, B and D, shows two atypical cells that responded considerably more vigorously to the phantom than to a single real source. As with the summed sinusoids, the spatial response profiles obtained with a phantom target was significantly broader than those obtained with a single speaker, when their variances were compared ($p < 0.001$; $n = 24$).

In contrast, when the two lateral speakers emitted random noise bursts generated from two, statistically uncorrelated random-number lists ($r = 0.005$), the space-specific neurons were able to resolve two distinct speakers. Results are shown in the right column of Figure 5A–D (open circles) for the same four

Figure 3. Spatial response profiles for four neurons (A–D) recorded in the ICx in response to stimulation with a sum of 10 sinusoids. Spike rate is plotted against the position of the center of the array of speakers for various array positions. Profiles shown in the *left column* were obtained in response to stimulation from a single speaker located in the center of the array. Thus, for these profiles, array position corresponds directly with speaker position. The *center* and *right columns* show responses to simultaneous stimulation from the two peripheral speakers. For these plots, the active speakers are approximately 22.5° to either side of the array position. The *center column* shows responses when all 10 frequencies are common to both speakers (*solid circles*), or unique to each speaker (*open circles*), or when each speaker emits five common and five unique frequencies (*triangles*). All frequencies have the same level. The *right column* shows responses when all 10 frequencies are common to both speakers but each frequency has a differing ratio of levels between speakers. A phantom source is thus generated in a different location for each frequency: clustered in the center (*solid circles*), split to either side (*open circles*), or spread evenly across the array (*triangles*).

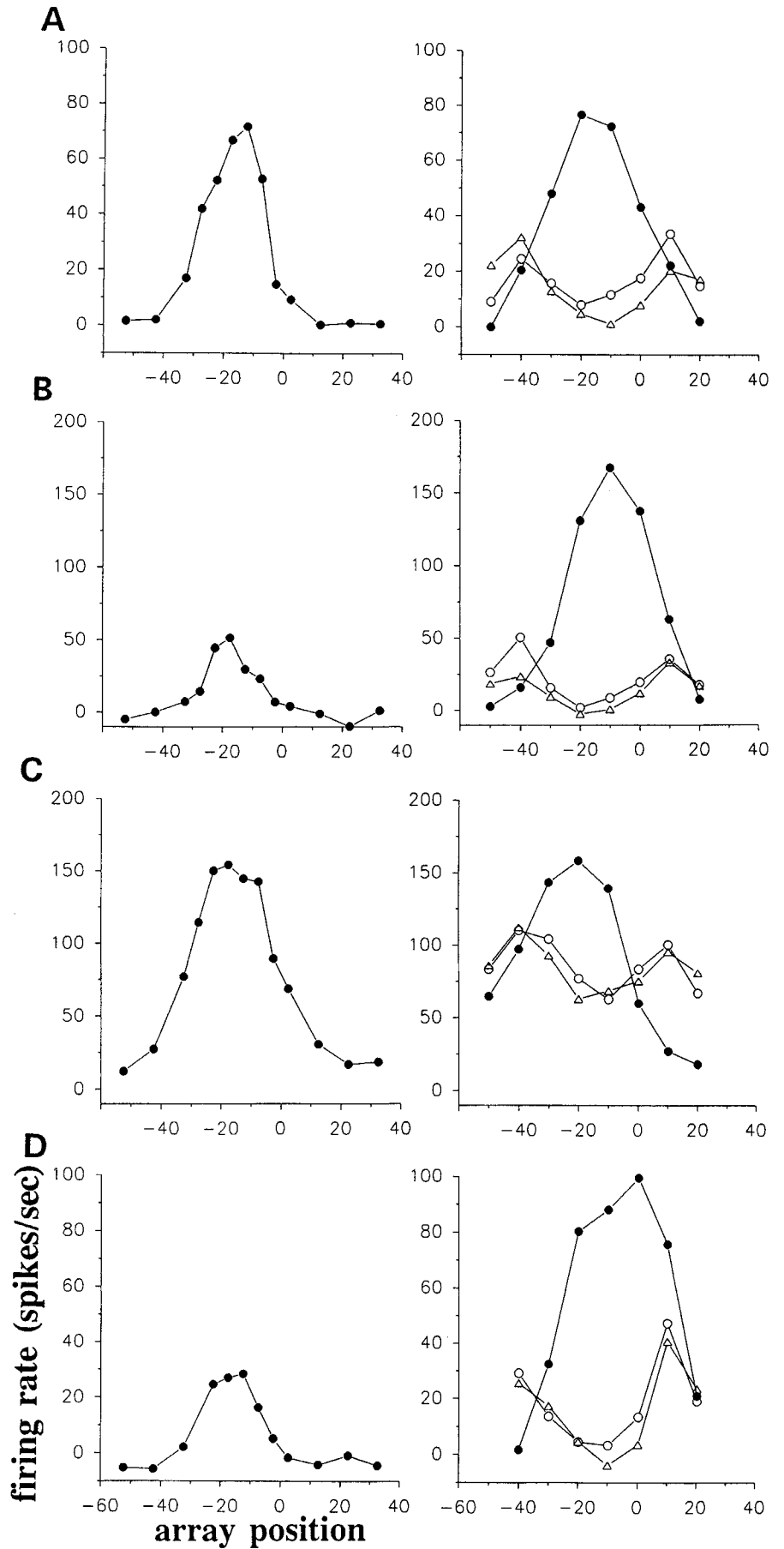


Figure 5. Spatial response profiles obtained with noise bursts for four neurons (A–D) recorded in the ICx. The *left column* shows profiles generated in response to stimulation from a single speaker. Thus, array position and speaker position are synonymous. In the *right column*, only the two lateral speakers, located 22.5° to either side of the array center, are simultaneously active. Three conditions are shown in the *right column*: identical noise bursts from each speaker (*solid circles*), temporally uncorrelated noise bursts from either speaker (*open circles*), and the same noise bursts from each speaker with the exception that one is temporally the reverse of the other (*triangles*).

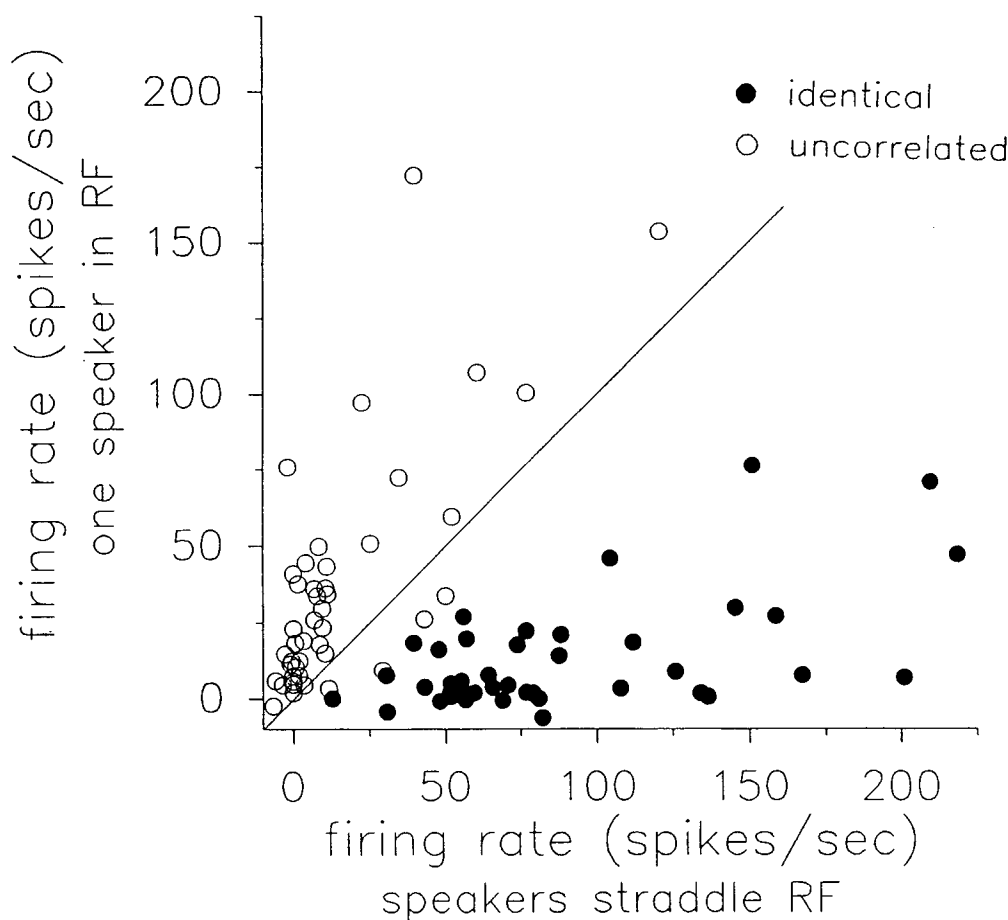


Figure 6. Response of 41 neurons recorded in the ICx when stimulated with one of two active speakers centered on the receptive field (*ordinate*) and when the two active speakers straddle the receptive field (*abscissa*). Each neuron was tested both with identical (*solid circles*) and uncorrelated noise bursts (*open circles*).

cells whose responses to identical noise bursts were just described. Each spatial response profile showed a distinct peak at each of the speaker locations and no phantom source was apparent. As with the profiles obtained with sounds composed of 10 different sinusoids, each of these peaks was significantly lower in magnitude than the single peak for either of the individual speakers or for a single phantom source ($p < 0.0001$; $n = 44$).

The difference in the responses to identical and uncorrelated noise bursts is quite consistent over our sample of cells ($p < 0.0001$; $n = 44$ cells). This is illustrated in Figure 6, a plot of each cell's response to one of the speakers in the receptive field as a function of their response to the two speakers straddling the receptive field. When the two speakers' noise bursts were statistically uncorrelated, the cells discharged more vigorously with one speaker in their receptive field (open circles). In contrast, when the noises were identical, the firing was stronger when the two speakers straddled the receptive field (solid circles).

The two statistically independent noise bursts used above have very similar spectra (though not identical), if those spectra are computed over the entirety of the bursts. However, if their spectra are compared to one another on a moment-by-moment basis, they will differ considerably more. Thus, the results above suggest that the space-specific neurons are exploiting the differences that occur on short time scales. To confirm this hypothesis we synthesized the two speakers' noises so that their long-term spectra were identical but their shorter-term spectra differed. This was done simply by producing a broadband noise burst from one speaker and simultaneously producing the same noise

burst from the other speaker, reversed in time. Spatial response profiles produced in response to these stimuli are shown by the triangles in the right column of Figure 5A–D). They were indistinguishable from those described above for uncorrelated noise bursts and showed a separate peak of activity for each speaker and a minimum of activity when the two speakers straddled the receptive field. These neurons could thus resolve two sound sources solely on the basis of differences in short-term spectra.

Discussion

The ability to parse the auditory scene depends heavily on localizing each source (Cherry, 1953; Sayers and Cherry, 1957; Bregman, 1990). This process is complicated by the fact that in nature, multiple objects emit sounds simultaneously, and the ears receive the sum of the sound waves from these simultaneously active sources. The owl's auditory system, which contains a topographic representation of space generated by the computation of binaural cues, has given us the chance to explore some basic principles by which an auditory system is able to resolve multiple, simultaneous sources. We demonstrated above that if two speakers produced identical summed-sinusoids, the space-specific neurons behaved as though they detected a single phantom source between the two speakers. On the other hand, if the two speakers' sounds were composed of sinusoids of different frequencies, space-specific neurons were able to resolve two separate targets. Our results also show that the position, size, and number of phantom targets can be controlled by manipulating the amplitude at which two sources emitted common frequencies. Finally, our tests show that space-specific neurons

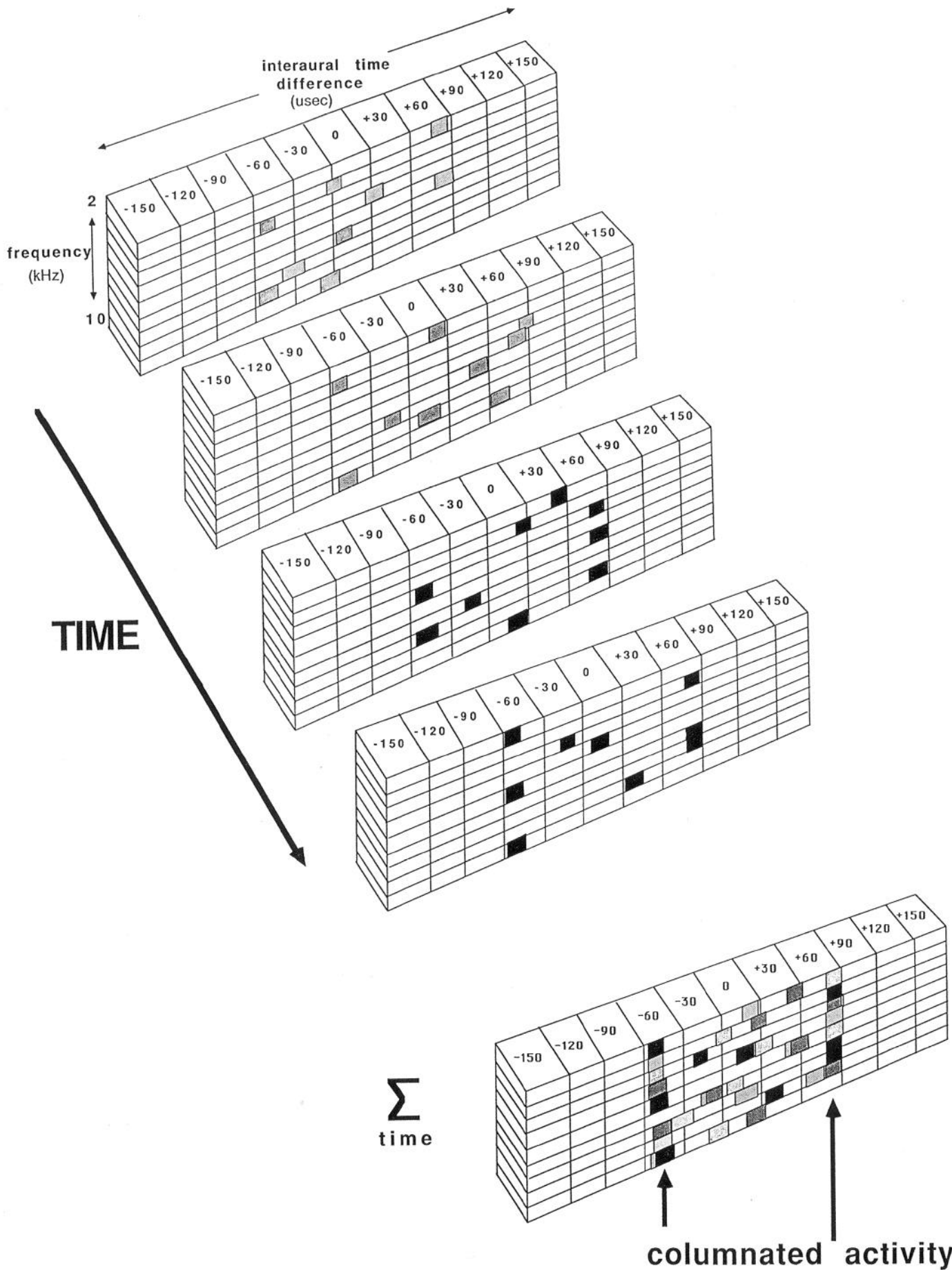


Figure 7. A model allowing representation of two sound sources in the ICx when presented with temporally uncorrelated noise bursts. See Discussion.

are capable of resolving two sources that emit statistically uncorrelated noise bursts with identical overall spectra, indicating that differences and similarities in spectra must be defined in the context of time.

Our data also suggest that multiple sources tax the ability of the space map to image targets. Even when neurons were capable of resolving two sources, the activity produced within the space map, as inferred from the spatial response profiles, was less distinct than that generated by a single source. A similar observation was reported by Mazer (1989), who studied the response of space-specific neurons to dichotically simulated multiple sound sources. This raises the question whether, behaviorally, the localization of a single source amid other active sources is correspondingly degraded. One simple reason for the neurophysiological observations may be that the space-specific neurons activated by each of the two sources were within the inhibitory surrounds of one another (Knudsen and Konishi, 1978b; Carr et al., 1989; Fujita and Konishi, 1991). Since the strength of lateral inhibition decreases as a function of distance across the map, we can expect that increasing the interspeaker distance will increase the magnitude and distinctness of the peaks in the spatial response profiles obtained with two sources.

Response to phantom sources

The response of space-specific neurons to phantom sources could be explained by the manner in which sound waves add. This, in turn, suggests that stereophonically generated phantom images are a flexible and simple method for manipulating the spatial attributes of auditory stimuli. Using the principle by which increasing the overall amplitude in one speaker biases the position of the phantom source toward that speaker, Grantham (1986) simulated moving sources and analyzed sensitivity to acoustic motion in humans. We have also used this method to confirm that some space-specific neurons are sensitive to the direction of stimulus motion (Wagner and Takahashi, 1990; unpublished observation). In our present experiments, this principle enabled us to generate conditions equivalent to two sources by using two sets of common frequencies and assigning opposite and relatively extreme interspeaker amplitude differences to them. Also, a single diffuse source was generated by using a wide range of interspeaker amplitude differences. The latter condition may prove to be particularly useful as it will allow us to generate phantom sources having various spatial extents.

Spectral constraints

We found that space-specific neurons can resolve two sources if the sounds they emit are composed of dissimilar frequencies. How dissimilar must the sounds be? Although critical bandwidths have not been measured for the barn owl, Quine and Konishi (1974) have estimated them to be between 1200 Hz (at 3500 Hz center frequency) and 2500 Hz (at 7000 and 10,000 Hz) from measurements of the Weber fraction. Furthermore, the cells in the owl's cochlear nuclei have Q_{10dB} values ranging from 2.3 to 8.75 (Sullivan and Konishi, 1984). Both of these measures of frequency resolution are broader than the interspeaker frequency separations (265–373 Hz; compare columns A and C of Table 1) used in the 10-unique stimulus, which the neurons were able to resolve (Figs. 3, 4). It is therefore likely that at least some frequencies from one speaker shared the tuning curve of a cochlear nucleus neuron with at least one frequency from the other speaker. The fact that the space-specific neurons were able to resolve two sources emitting these rela-

tively closely spaced frequencies suggests that higher auditory nuclei can resolve frequencies that fall within the tuning curve of neurons in the cochlear nuclei.

The superposition of spectral components points out clearly that binaural information in single frequency channels is unreliable, for it can represent a single source, or it can represent the weighted average of the binaural cues from multiple sources. The accurate parsing of an auditory environment containing simultaneously active sources thus requires access to information in multiple frequency channels. It is generally agreed that sound localization requires access to broadband signals because binaural cues are more richly represented and because access to multiple frequency channels resolves ambiguities associated with a single channel (Yin and Kuwada, 1983; Takahashi and Konishi, 1986; Yin et al., 1987; Stern et al., 1988). The requirements for the representation of multiple, concurrent sounds provide yet another compelling reason why cross-frequency analysis is necessary.

Temporal integration and resolving multiple noise sources

Broadband noises are of particular ethological importance to the barn owl as they are the sounds made by its natural prey (Payne, 1971; Konishi, 1973). Our results show that space-specific neurons are indeed able to resolve two sources emitting statistically uncorrelated broadband noises or two sources emitting broadband noises that are temporally reversed versions of each other. Similar results were also reported in a psychoacoustical study by Damaske (1967/1968), in which human listeners were asked to describe the extent of the sound field when two loudspeakers simultaneously emitted broadband noise. The subjects indicated two distinctive sources when the two speakers' noises were uncorrelated. These sources fused together, if the speakers' sounds were correlated, or if the speakers were brought closer together. Knudsen and Konishi's (1978b) demonstration of center-surround antagonism in the receptive fields of space-specific neurons indirectly showed that the space map can resolve two uncorrelated noise sources. They showed that a space-specific neuron's response to noise bursts from a speaker in its receptive field could be inhibited by a second speaker, located up to 70° away, that emitted noise bursts from an independent noise generator. For this to have worked, the two speakers must have retained their spatial identities on the auditory space map.

Since the spectra of the two speakers' sounds in our experiments, computed over their entirety, are identical or nearly so, our results suggest that space-specific neurons exploit differences in spectra that arise over shorter time spans. How might the space-specific neurons utilize these short-term spectral differences? When noise bursts are used, the pattern of electrical activity in ICc is likely to be dynamic and complex. If the noises emitted have different short-term spectra, common and unique frequencies will come and go from moment to moment. In ICc, the activity evoked by unique frequencies will always be in columns that correspond to the sources' actual Δt values. Activity evoked by spectral components that are too similar to be resolved would be expected to move between columns, first, because the relative contribution of the two noise sources to any given frequency can be expected to change over time, and also because of the way in which frequencies might add together at more peripheral levels, as discussed above. Such a scenario is depicted in Figure 7, which shows the hypothetical activity in ICc at four consecutive instants in time. Note that at any instant,

distinctive columns are not visible in ICc. However, if the activity in ICc is summed over time, activity evoked by the unique frequencies will be reinforced in the columns that represent the sources' true Δt values, whereas the activity evoked by common frequencies will be scattered without generating a columnar focus. Thus, if ICx were to compute such a temporal sum, space-specific neurons could be activated strongly by the columnated activity. The scattered activity, by contrast, will activate space-specific neurons only weakly. Furthermore, it is conceivable that these weakly driven space-specific neurons will be suppressed by the strongly driven space-specific neurons, through lateral inhibitory connections (Knudsen and Konishi, 1978b; Carr et al., 1989; Fujita and Konishi, 1991). The precise time over which the space-map might compute this sum has yet to be measured, but Wagner (1992) has shown that space-specific neurons require about 2 msec to detect binaural correlation. It would not be surprising if a similar amount of time were required for the space-specific neurons to resolve two simultaneous sound sources.

The scenario presented above requires the inputs to ICx to deal with the potentially conflicting demands of time and frequency resolution. At any given instant, the activity in ICc reflects the spectrum of sounds in the environment averaged over some period of time. In order to take advantage of short-term differences in the spectra of the two source's sounds, it is best to keep the averaging time short and to update its activity pattern rapidly. However, if the inputs leading to ICx behaved as linear filters, their abilities to respond to rapid changes in frequencies would necessarily degrade their abilities to resolve frequencies. The ability to parse the two noise bursts that differ only in their temporal fine structure would seem to place rigorous demands on both temporal and spectral resolution. It would not be surprising, therefore, if the brainstem auditory pathways had evolved some nonlinear strategy allowing high resolution in both domains.

From "where" to "what" in the auditory system

The columnar organization of the ICc sorts frequencies according to their azimuth. In this way, it creates auditory "objects" defined as a spectrum originating from a discrete location. Our study suggests that when there are multiple sources in the environment, strong columnar activity requires frequencies that are unique to each source. Frequencies that cannot be resolved produce activity that cannot be localized to a column. This suggests several questions: does the auditory system make use of frequency information that cannot be assigned to a given source in space? And, if an object is defined by a column in ICc, does this mean that the complete spectrum of an object cannot be determined in the presence of other sound sources that have similar spectral components? A study by Divenyi and Bregman is particularly relevant (as cited in Bregman, 1990; also Divenyi, personal communication). These investigators asked human listeners to localize sound sources when two speakers emitted sounds in which all spectral components except one were unique to each source. The one common component was emitted from the two sources at equal amplitudes. The subjects did not perceive three sources, that is, the two speakers emitting unique frequencies from the sides and a phantom source representing the common frequency in the middle, as one might expect from considering only the physical signal and binaural processing. Instead, these subjects perceived two sources and attributed the common component to both of the sources. Thus, the human

auditory system seems to be able to reallocate common frequencies to sources that have distinct spatial locations. This process would accurately determine not only the number of sources, but also the complete spectrum of each source as well. At present, there is no evidence to suggest that the space map of the owl is capable of a similar process.

References

- Bauer BB (1961) Phasor analysis of some stereophonic phenomenon. *J Acoust Soc Am* 33:1536–1539.
- Blauert J (1983) *Spatial hearing*. Cambridge, MA: MIT Press.
- Bregman AS (1990) *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Carr CE, Konishi M (1990) A circuit of detection of interaural time differences in the brain stem of the barn owl. *J Neurosci* 10:3227–3246.
- Carr CE, Fujita I, Konishi M (1989) Distribution of GABAergic neurons and terminals in the auditory system of the barn owl. *J Comp Neurol* 286:190–207.
- Cherry EC (1953) Some experiments on the recognition of speech with one and with two ears. *J Acoust Soc Am* 25:975–979.
- Damaske P (1967/1968) Subjektive Untersuchungen von Schallfeldern. *Acustica* 19:198–213.
- Fujita I, Konishi M (1991) The role of GABAergic inhibition in processing of interaural time difference in the owl's auditory system. *J Neurosci* 11:722–739.
- Geisler CD, Rhode WS, Hazelton DW (1969) Responses of inferior colliculus neurons in the cat to binaural acoustic stimuli having wide band spectra. *J Neurophysiol* 32:960–974.
- Goldberg JM, Brown PB (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* 32:613–636.
- Grantham DW (1986) Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J Acoust Soc Am* 79:1939–1949.
- Knudsen EI (1983) Subdivisions of the inferior colliculus in the barn owl (*Tyto alba*). *J Comp Neurol* 218:174–186.
- Knudsen EI, Konishi M (1978a) Space and frequency are represented separately in the auditory midbrain of the owl. *J Neurophysiol* 41:870–884.
- Knudsen EI, Konishi M (1978b) Center surround organization of auditory receptive fields in the owl. *Science* 202:778–780.
- Knudsen EI, Konishi M (1979) Mechanisms of sound localization in the barn owl (*Tyto alba*). *J Comp Physiol* 133:13–21.
- Konishi M (1973) Locatable and nonlocatable acoustic signals for barn owls. *Am Nat* 107:775–785.
- Konishi M, Takahashi TI, Wagner H, Sullivan WE, Carr CE (1988) Neurophysiological and anatomical substrates of sound localization in the owl. In: *Auditory function: development, physiology, psychophysics* (Edelman GM, Gall WE, Cowan WM, eds), pp 721–746. New York: Wiley.
- Mazer JA (1989) Single unit responses to multiple sound sources in the inferior colliculus of the owl. *Soc Neurosci Abstr* 15:115.
- Moiseff A (1989) Binaural disparity cues available to the barn owl for sound localization. *J Comp Physiol* 164:629–636.
- Moiseff A, Konishi M (1981) Neuronal and behavioral sensitivity to binaural time differences in the owl. *J Neurosci* 1:40–48.
- Moiseff A, Konishi M (1983) Binaural characteristics of units in the owl's brainstem auditory pathways: precursors of restricted spatial receptive fields. *J Neurosci* 3:2553–2562.
- Payne RS (1971) Acoustic location of prey by barn owls (*Tyto alba*). *J Exp Biol* 54:535–573.
- Quine DB, Konishi M (1974) Absolute frequency discrimination in the barn owl. *J Comp Physiol* 93:347–360.
- Rose JE, Gross NB, Geisler CD, Hind JE (1966) Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. *J Neurophysiol* 29:288–314.
- Sayers BM, Cherry EC (1957) Mechanism of binaural fusion in the hearing of speech. *J Acoust Soc Am* 29:973–987.
- Stern RM, Zeiberg AS, Trahiotis C (1988) Lateralization of complex binaural stimuli: a weighted-average model. *J Acoust Soc Am* 84:156–165.
- Sullivan WE, Konishi M (1984) Segregation of stimulus phase and

- intensity coding in the cochlear nucleus of the barn owl. *J Neurosci* 4:1787-1799.
- Sullivan WE, Konishi M (1986) Neural map of interaural phase difference in the owl's brainstem. *Proc Natl Acad Sci USA* 83:8400-8404.
- Takahashi TT, Konishi M (1986) Selectivity for interaural time difference in the owl's midbrain. *J Neurosci* 6:3413-3422.
- Takahashi TT, Konishi M (1988) Projections of the cochlear nuclei and nucleus laminaris to the inferior colliculus of the barn owl. *J Comp Neurol* 274:190-211.
- Takahashi TT, Moiseff A, Konishi M (1984) Time and intensity are independently processed in the auditory system of the owl. *J Neurosci* 4:1781-1786.
- Takahashi TT, Wagner H, Konishi M (1989) The role of commissural projections in the representation of bilateral auditory space in the barn owl's inferior colliculus. *J Comp Neurol* 281:545-554.
- Wagner H (1992) On the ability of neurons in the barn owl's inferior colliculus to sense brief appearances of interaural time difference. *J Comp Physiol [A]* 170:3-11.
- Wagner H, Takahashi TT (1990) Neurons in the midbrain of the barn owl are sensitive to the direction of apparent acoustic motion. *Naturwissenschaften* 77:439-442.
- Wagner H, Takahashi TT, Konishi M (1987) Representation of interaural time differences in the central nucleus of the barn owl's inferior colliculus. *J Neurosci* 7:3105-3116.
- Yin TCT, Chan JCK (1989) Neural mechanisms underlying interaural time sensitivity to tones and noise. In: *Auditory function: development, physiology, psychophysics* (Edelman GM, Gall WE, Cowan WM, eds), pp 385-430. New York: Wiley.
- Yin TCT, Chan JCK (1990) Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol* 64:465-488.
- Yin TCT, Kuwada S (1983) Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency. *J Neurophysiol* 50:1020-1042.
- Yin TCT, Chan JCK, Carney LH (1987) Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. III. Evidence for cross-correlation. *J Neurophysiol* 58:562-583.