Binaural Cross-Correlation Predicts the Responses of Neurons in the Owl's Auditory Space Map under Conditions Simulating Summing Localization

Clifford H. Keller and Terry T. Takahashi

Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403-1254

Summing localization describes the perceptions of human listeners to two identical sounds from different locations presented with delays of 0–1 msec. Usually a single source is perceived to be located between the two actual source locations, biased toward the earlier source. We studied neuronal responses within the space map of the barn owl to sounds presented with this same paradigm. The owl's primary cue for localization along the azimuth, interaural time difference (ITD), is based on a cross-correlation-like treatment of the signals arriving at each ear. The output of this cross-correlation is displayed as neural activity across the auditory space map in the external nucleus of the owl's inferior colliculus. Because the ear input signals reflect the physical summing of the signals generated by each speaker, we first recorded the sounds at each ear and computed their cross-correlations at various inter-

stimulus delays. The resulting binaural cross-correlation surface strongly resembles the pattern of activity across the space map inferred from recordings of single space-specific neurons. Four peaks are observed in the cross-correlation surface for any nonzero delay. One peak occurs at the correlation delay equal to the ITD of each speaker. Two additional peaks reflect "phantom sources" occurring at correlation delays that match the signal of the left speaker in one ear with the signal of the right speaker in the other ear. At zero delay, the two phantom peaks coincide. The surface features are complicated further by the interactions of the various correlation peaks.

Key words: auditory scene analysis; echo suppression; inferior colliculus; interaural time difference; precedence effect; sound localization

In nature, sounds arriving directly from an active source are often overlapped with echoes, affecting the cues by which we perceive auditory space. To understand how echoes affect spatial hearing, we have examined the responses of neurons in the barn owl's (*Tyto alba*) map of auditory space to a direct sound followed shortly thereafter by a simulated echo. The owl's space map consists of an array of neurons, called space-specific neurons, that are selective for binaural cues and therefore have spatial receptive fields. The pattern of activity across the space map identifies the sound sources available to the owl for localization.

We demonstrated previously (Keller and Takahashi, 1996) that when an echo follows the direct sound by 0.5–5.0 msec, the response of a space-specific neuron to the echo is suppressed, suggesting that the image of the echo on the space map is weakened. This parallels the phenomenon of the precedence effect in which human subjects localize only the first of two sound sources activated in rapid sequence (Wallach et al., 1949; Haas, 1951). Below, we examine the responses of neurons to shorter delays, which in humans give rise to a different phenomenon called summing localization. In summing localization, subjects generally report a single sound source that seems to be at a position between the two sources, biased toward the leading source. Experienced listeners may report additional sources. Sum-

ming localization is experienced only for highly correlated sounds (Damaske, 1969/70), suggesting that the nature of the signals at the ears may offer clues to explain the perceptual effects.

One of the primary cues for the localization of sounds is the ongoing interaural time difference (ITD) in the arrival time of sounds. ITD is generally thought to be derived by cross-correlating the signals of the two ears (Sayers and Cherry, 1957; Stern et al., 1988). According to the model of Jeffress (1948), action potentials from the cochlear nuclei of both sides, which are phase-locked to a particular spectral component, converge on a neuron that discharges maximally when the inputs from the two sides arrive simultaneously. The phase-locked action potentials are delayed on one side by the ITD so that the coincidence occurs within the postsynaptic nucleus only where the axonal lengths impose a delay that compensates for the ITD.

Extensive evidence for the cross-correlation model of Jeffress has accrued in the mammalian auditory system (Rose et al., 1966; Geisler et al., 1969; Goldberg and Brown, 1969; Kuwada and Yin, 1983; Yin and Kuwada, 1984; Yin et al., 1987; Yin and Chan, 1988, 1990). In the owl, the function of the delay lines is subserved by the axons of the nucleus magnocellularis, and the role of the coincidence detectors is filled by the neurons of the nucleus laminaris (Sullivan and Konishi, 1984, 1986; Carr and Konishi, 1990). Nucleus laminaris projects to the inferior colliculus (ICx) where information from multiple frequency channels is combined to derive a topographic map of azimuth in ICx (Wagner et al., 1987). Recent evidence indicates that space-specific neurons within the ICx are sensitive to the level of correlation of the signals of the two ears (Albeck and Konishi, 1995).

Given the central role of cross-correlation in spatial hearing, we first describe the binaural cross-correlation of signals re-

Received Jan. 29, 1996; revised April 8, 1996; accepted April 11, 1996.

This research was supported by grants from the Whitehall Foundation and National Institute of Deafness and Communication Disorders. We thank Drs. T. C. T. Yin and Petr Janata for helpful discussions and criticisms, and Petr Janata for technical assistance.

Correspondence should be addressed to Dr. Clifford H. Keller, Institute of Neuroscience, 222 Huestis Hall, University of Oregon, Eugene, OR 97403-1254.

Copyright © 1996 Society for Neuroscience 0270-6474/96/164300-10\$05.00/0

Figure 1. *A*, Schematic representations of the ear input signals resulting when two loudspeakers, separated in azimuth, emit identical sounds with a slight delay (*ISI*). For clarity, identical portions of the sound arriving from each source are shown. The sound from the left loudspeaker is shown as a *solid line*, that from the right loudspeaker is a *dashed line*. The actual ear input signals would comprise a mixture of these signals and those from any other sources. Speaker locations are shown in the drawing to the left. ITD_L and ITD_R signify the interaural time differences corresponding to the azimuthal location of the left and right speaker, respectively. *B*, Cross-correlation calculated from a broad-band signal such as that shown in *A*. Four peaks occur at correlation delays (τ) corresponding to $\pm$ ISI, ITD_L , and ITD_R .

coded in the ear canals when two sources separated in azimuth are activated with short delays ranging up to several hundred microseconds. The cross-correlations obtained at these short delays are then compared to the responses of neurons in the auditory space map.

MATERIALS AND METHODS

Neurophysiological recordings were obtained from five captive-bred, adult barn owls. Anesthetic and surgical procedures for neurophysiological recordings, which have been published previously (Takahashi and Keller, 1994), were approved by the institutional animal care and use committee of the University of Oregon. Briefly, an owl was anesthetized with ketamine (100 mg/ml Vetalar, Parke-Davis; 0.1 ml, i.m., approximately every 2 hr) and diazepam (5 mg/ml Diazepam, C-IV, LyphoMed; 0.05 ml, i.m., approximately every 2 hr) and held within a stereotaxic device by a stainless steel plate cemented to the skull. All recordings were carried out within an echo-attenuating booth (Industrial Acoustics; 1.8 m $\times$ 1.8 m $\times$ 1.8 m inner dimension, lined with 15.2 cm Ilbruck Sonex acoustic foam). The responses of single space-specific neurons were recorded using epoxy-insulated tungsten microelectrodes (Fredrick Haer, 10 M Ω). Action potentials were amplified and level-discriminated, and the times of their occurrence relative to stimulus onset were written to a computer file. Stimuli consisted of 100 msec bursts of broad-band noise flat within ± 2 dB between 2,000 and 9,000 Hz after transduction. The noise was synthesized digitally with 12-bit resolution, converted to analog form at 50,000 samples/sec (Modular Instruments), and multiplied by a trapezoidal envelope (5 msec onset, 5 msec offset). Sounds were then amplified (McIntosh, M754) and attenuated (Tucker Davis Technologies, PA4) to produce sound pressure levels 20–30 dB above neuronal thresholds.

When a space-specific neuron was isolated, the receptive field of the cell was evaluated with 5° spatial resolution by plotting the number of action potentials as a function of the azimuth of a 2-cm-diameter speaker (Alpine 6020HX). The speaker was mounted at eye level on a semicircular hoop that could be pivoted about an imaginary vertical line through the center of the owl's head at the anteroposterior level of the ear openings. To record the response of the cell under conditions of summing localization, this procedure was repeated using two speakers (Alpine 6020HX), spaced apart by 45 or 55° of azimuth, mounted on the same hoop. Each speaker emitted identical noise bursts with delays ranging from -500 μ sec (left speaker leading) to $+500$ μ sec (right speaker leading).

Ear canal recordings were obtained from two owls using the acoustical

stimulus paradigm described above. Because our goal was to compare the predictions from ear canal recordings with neuronal responses, we took care to replicate the conditions that are normally present during a neurophysiological experiment. Thus the owl was placed in the stereotaxic device within the sound-isolating booth used in neurophysiological experiments. Small microphones (Knowles EM 4046) were inserted as far into the ear canals as possible without risking damage to the tympanic membrane or its surrounding tissue. Typically, the microphone port was 5 mm from the tympanic membrane and facing outward. The microphones had matched frequency-response curves flat to within ± 6 dB between 3000 and 9000 Hz, the effective frequency range for sound localization in the owl (Knudsen and Konishi, 1979). The amplified output of the microphones was digitized by a computer-controlled analog-to-digital converter (Tucker Davis Technologies, PD1) at a rate of 100,000 samples/sec. Binaural cross-correlations were computed from a 40.96 msec segment of these digitized ear-canal recordings using the XCORR function of the MATLAB software package (version 4.2c.1, The MathWorks).

RESULTS

Binaural cues and the binaural cross-correlation

Figure 1 schematically depicts the signals arriving at the ears from two loudspeakers, separated in azimuth, that emit identical noise bursts with a slight delay (interstimulus interval, ISI). In Figure 1*A*, short portions of the arriving signals are plotted for each ear, relative to the time of arrival of the sound from the leading (left) source. Identical portions of the sound emitted from each loudspeaker (*solid lines* from left loudspeaker, *dashed lines* from right loudspeaker) arriving in each ear are shown. At the left ear (*top trace*), the identical waveform is received first from the leading left speaker (at 0 delay, as per our convention) and then from the right speaker at a delay equal to the ISI plus the ITD of the right speaker (ITD_R). The right ear (*bottom trace*) receives first the sound from the left (leading) speaker with a delay equal to the ITD of the left speaker (ITD_L) and then the sound from the right speaker with a delay equal to the ISI.

The cross-correlation of the two ear-input signals is shown in Figure 1*B*. Binaural cross-correlation first involves shifting the signal of one ear by a delay τ and performing a point-by-point

multiplication of the signals of the left and right ears. The products are summed and plotted as a function of τ , and the entire process is repeated for a range of τ . The sums, or correlation levels, reach maxima when the delay τ brings similar or identical segments of the signals of the left and right ears into alignment. For example, if the signal of the right ear is shifted to the left by an amount equal to ITD_L , the correlation level will reach a maximum. The same will hold true if $\tau = ITD_R$. Note also that when the signal of the right ear is shifted further to the left by an amount equal to the ISI, the contribution of the right speaker to the right ear is aligned with the contribution of the left speaker to the left ear. This will cause another maximum in the cross-correlation, the position of which depends on the ISI. Similarly, if the signal of the right ear is shifted further to the right by ISI, another maximum is created. The latter maxima are phantom targets and do not correspond to any actual sound source, and their heights depend on the similarity of the sounds coming from the two loudspeakers. τ is analogous to the ITD, and the neurons in the space map can be said to be selective for a narrow range of τ . In the space map, therefore, cross-correlation should give rise to at least four areas of strong neural activity representing the two speakers at ITD_L and ITD_R and the two phantom targets at $+ISI$ and $-ISI$.

To evaluate the binaural cross-correlation of sounds generated within our experimental conditions, we placed miniature microphones into the external ear canals of the owl and recorded the signals received in each ear when the bird was presented with noise bursts from speakers placed 27.5° to either side of the bird's midline at eye level. The speakers emitted identical, broad-band, 100 msec noise bursts with ISIs between $\pm 300 \mu\text{sec}$. We then cross-correlated these two ear-input signals at each ISI and plotted the results as a correlation surface (Fig. 2A). To show more clearly the structure of this surface, we have plotted the τ -axis for a range of $\pm 500 \mu\text{sec}$, which is roughly double the range of ITDs encountered by the barn owl in nature. The surface has been collapsed onto a schematized and enlarged planar view in Figure 2B (dashed lines).

The cross-correlation surface is dominated by features corresponding to each of the peaks seen previously in Figure 1B. Two parallel lines (dashed lines in Fig. 2A) reflect high binaural cross-correlations when $\tau = ITD_L$ or $\tau = ITD_R$, the positions of which do not change with the ISI. Because the two sources emit identical sounds, two other peaks of the correlation function correspond to phantom sources generated by the binaural fusion of sounds from the two separate speakers. Because the τ -values associated with these peaks are functions of and actually equivalent to the ISI, these peaks are seen as diagonal lines where $|\tau| = |ISI|$ (dotted lines in Fig. 2A).

Note that for much of the surface plotted in Figure 2, the phantom sources occur more peripherally than the two actual speaker locations. The two diagonals intersect at a central peak where both ISI and τ are roughly equal to zero and only a single, centrally located phantom is generated. In the schematized Figure 2B, even near 0 ISI, the two real sources should generate high levels of correlation and the parallel lines should remain unbroken. Note that in the actual microphone recordings of Figure 2A, however, because of the bandpassed nature of the sounds, each correlation feature is actually the peak of a highly damped oscillation along the τ -axis, and the details of the surface are complicated by the interaction of the peaks and valleys corresponding to each correlation feature.

Figure 2 suggests that considering only the outcome of cross-

Figure 2. Binaural cross-correlation surface generated when two loudspeakers placed 27.5° to either side of the owl's midline emit identical broad-band noise bursts with various delays (ISIs). *A*, The binaural cross-correlation is represented by a gray scale along the vertical axis (white, maximum positive correlation; black, maximum negative correlation) for a range of cross-correlation delays (τ) and interstimulus delays. This surface is shown within dashed lines as part of an expanded, schematic drawing in *B*. The schematic highlights three prominent features of the binaural cross-correlation: (1) two parallel lines that represent the interaural time difference corresponding to each of the actual speaker locations (dashed lines in *A*); (2) two diagonal lines that represent "phantom sources" (dotted lines in *A*), computed as the interaural time difference between the sound from one speaker and the (delayed) sound from the other speaker; and (3) the intersection of these diagonals where both the cross-correlation delay and the ISI = 0 (white arrow).

correlation-like processes, and given identical sounds emitted from each speaker, the sources available for localization should always include the two real sources and two phantom sources, except when $ISI \cong 0$ and only one phantom source occurs. If, on the other hand, the sounds from the two loudspeakers are uncorrelated, the binaural cross-correlation will not show peaks corresponding to phantom sources (Fig. 3). Only the two parallel ridges are seen, and only two sources should be localizable, each to its true location. When sounds from the two loudspeakers are partially correlated, phantom peaks of lower magnitude are seen.

Patterns across the space map

The binaural cross-correlation surfaces shown in Figures 2 and 3 are predictions for the output of correlation-like neural mecha-

Figure 3. Binaural cross-correlation surface resulting when the two noise bursts are uncorrelated. Presentation and speaker arrangement are the same as in Figure 2. Note the absence of the crossing diagonals resulting from phantom sources as seen in Figure 2.

nisms of the auditory system. In the barn owl, the output of nucleus laminaris is ultimately displayed as the activity of space-specific neurons, which are arrayed in the ICx to form a topographic map of auditory space. We wish to understand how activity is distributed across this display under conditions that simulate summing localization and to compare this distribution to the predictions of Figures 2 and 3. Although possible, it would be quite cumbersome to sample the activity of different neurons across the map while leaving two speakers at fixed locations. Instead, by assuming that all space-specific neurons respond similarly, regardless of the location of their receptive field, it is possible to infer the activity pattern of the space map by recording the response of a single space-specific neuron to stimuli that elicit summing localization with the speaker array located at various azimuths. In practice, this procedure is much like determining the receptive field of a cell, except that two speakers are used instead of one. An additional assumption of this method is that only the spatial location of the sound is being changed. It is quite clear, however, that the filter characteristics of the ears vary with the spatial location of the sound and thus the auditory scene computed from the binaural cues at each array location may differ. Thus, to compare most rigorously the predicted activity patterns generated by binaural cross-correlation with the responses of space-specific neurons, we computed the cross-correlations from ear-input signals gathered with the speaker array located at the same azimuths as for neurophysiological recording. Figure 4 illustrates this procedure.

We wish to ascertain the activity across the space map when two loudspeakers are located 27.5° to either side of the midline, the same conditions used in constructing Figures 2 and 3. Consider the responses of a hypothetical cell that is narrowly tuned to spatial locations directly in front of the owl (Fig. 4A; RF and large arrow). Having determined the best azimuth of the cell, we centered the loudspeaker array 60° to the left of the best azimuth (at -60° because the best azimuth was 0°) and recorded the firing of this cell when presented with various ISIs (abscissa). This situation is analogous to centering the speaker array at 0° and recording from a cell whose receptive field was centered 60° to the right of the midline ($+60^\circ$, dashed arrow in Fig. 4A). We therefore assigned this activity to $+60^\circ$ of azimuth along the abscissa of

Figure 4. Spatial correlation surface computed for identical sounds emitted by two speakers located 27.5° to either side of the midline. The speaker array was centered at positions between $\pm 90^\circ$ azimuth in 5° increments and at each array location, and for each ISI, the value of the binaural cross-correlation for $\tau = 0$ is plotted. Two parallel lines of high correlation correspond to the speaker locations at approximately $\pm 27.5^\circ$. A centrally located phantom source can be seen near 0 ISI. Note the apparent absence of any peaks in the correlation that would correspond to phantom sources at azimuths extreme to the actual speaker locations.

Figure 4B. We repeated this process with the speaker array at various azimuths to infer the pattern of activity across the entire (bilateral) space map while recording from only a single cell.

We can analogously derive an entire map of the binaural cross-correlation function as it would have been computed by the hypothetical cell. We compute the binaural cross-correlation at each array location and extract the correlation level at the τ (or ITD) to which the hypothetical cell is maximally responsive (0 μsec for our hypothetical cell in Fig. 4). This value is then plotted in the same manner as was the firing rate of the cell, and a map of the cross-correlation function is obtained (Fig. 4B). Such plots are termed “spatial response surfaces” when referring to the inferred activity of neurons across the map and “spatial correlation surfaces” when referring to the analogously derived binaural cross-correlation surface.

For the example above, we used a narrowly tuned cell and extracted the correlation value at τ equivalent to the best azimuth of the cell. Many cells, however, are tuned more broadly. To predict the responses of these cells, we weighted (see figure legends) and summed correlation values over a range of τ to reflect the single-speaker spatial tuning characteristics of the cell.

Spatial correlation surfaces obtained in this manner show strong peaks that correspond to ITD_L and ITD_R . Relatively weaker peaks correspond to phantom sources, and these peaks coalesce into one central peak near 0 ISI. Each of these features shows strong modulation over the range of ISIs tested as the peaks and troughs of each feature interact along the τ -axis. In contrast to the correlation surface of Figure 2A, however, the phantom sources do not extend peripheral to the real sources. This is probably attributable to both the interactions of the various correlation features and the fact that binaural correlations weaken markedly as even single sources are placed more laterally (see below).

Responses of neurons under conditions that may elicit summing localization

We recorded from 47 individual space-specific neurons in five owls. We presented each cell with a range of ISIs and mapped their responses over a range of speaker-array locations. We compared these responses with spatial correlation surfaces that predict the pattern of activity across the space map when two loudspeakers were arranged to straddle the midline. These comparisons for all cells showed similar patterns, which are exemplified by the responses described below.

The cross-correlation plots of Figures 2 and 4 show that identical sounds presented simultaneously from two speakers create a phantom source halfway between the two speakers. This phantom source splits in two and moves toward either side as the magnitude of the ISI is increased. Thus, if the speakers were located to either side of the receptive field of a cell, a response only for ISIs near 0 μ sec would be expected. Figure 5 shows the responses of a space-specific neuron, the receptive field of which was centered at -5° , when the cell was presented with identical noise bursts with various ISIs. Raster plots giving the times of action potentials are shown to the left. In Figure 5A, the loudspeakers straddle the receptive field and a strong response is seen at ISIs near 0 μ sec. Weaker responses are also seen over a range of ISIs from -120 to -180μ sec and $+120$ to $+180 \mu$ sec. The plots to the right show the spike rate of the cell at each ISI, normalized to the maximum spike rate over all tests (*filled circles*, left ordinate axis). These spike rates can be compared with the values of the binaural cross-correlation at $\tau = -10 \mu$ sec (approximately -5°), with the speaker array in the same location (*open circles*, right ordinate axis). The overall shapes of the two plots are quite similar, and although the ordinate scales for the two plots are not directly comparable, the data also suggest that the signal-to-noise ratio of the cellular response might be greater than for the binaural cross-correlation. The plots of Figure 5B,C present the responses and binaural cross-correlations for the same cell when the speaker array was centered at $+10^\circ$ and $+25^\circ$, respectively. These plots show a strong modulation of the response as the ISI is changed, even when the left loudspeaker is located directly within the center of the receptive field of the cell. This modulation results from the strong interaction along the τ -axis between the various correlation features, and these interactions depend on the band-passed nature of the sound and the filtering properties of the external auditory apparatus. Comparison of the spike rate curves with the associated binaural cross-correlation functions shows a strong match for these array locations as well.

Figure 6 allows comparison of the entire spatial correlation surface with the spatial response surface of the same cell as in Figure 5. The correlation values (Fig. 6A) were extracted as the weighted sum of values centered at $\tau = -10 \mu$ sec (approximately

-5°) to approximate the single-speaker spatial tuning curve of the cell, which is shown in Figure 6C. The similarities of the two surfaces are striking. For example, there is a strong response to the left speaker at approximately -30° and to the right speaker at approximately $+30^\circ$ at any ISI. The response to the left speaker, however, shows peaks between -150 and -100μ sec ISI and between $+10$ and $+50 \mu$ sec ISI. At these ISIs where there are peaks in the response to the left speaker, there are troughs in the response to the right speaker, and vice versa. A relatively weak and broad central phantom is seen at ISIs near 0 μ sec. This phantom shows a weak diagonal trend from upper left to lower right. There are no responses to phantom sources peripheral to the actual speaker locations. All of these features are seen in both the response of the cell and the spatial correlation surface.

Similar surfaces are shown for two more cells in Figures 7 and 8. The cell in Figure 7 has a relatively broad receptive field, centered at approximately -25° (equal to -57.5μ sec ITD at 2.3μ sec/degree; Moiseff, 1989). The response of this cell is compared with the correlation values centered at a τ of -60μ sec. The cell shown in Figure 8 had a receptive field centered at -10° (-23μ sec ITD) and is compared with correlation values centered at $\tau = -20 \mu$ sec. In each case, the responses of the cells closely match those predicted by the spatial correlation surfaces. The most prominent features are the two deeply modulated parallel lines that correspond to actual speaker locations. The depths of modulations and the ISIs where peaks and troughs are found depend on the receptive field of the cell, but they match well with the patterns predicted by the correlation surfaces. The interactions between correlation features are strongest at ISIs between approximately $\pm 200 \mu$ sec, resulting in an alternation of peak correlation levels between the two parallel lines as the ISI is changed. At longer ISIs, the correlation strengths allied with each speaker are more consistent and equal (not shown). These same patterns are seen in the neural responses until at ISIs greater than approximately $\pm 500 \mu$ sec a suppression of the lagging source is seen (Keller and Takahashi, 1996). The threshold ISI at which this suppression takes effect has not been explored thoroughly. In each case there is also a diagonally extending phantom source that crosses from ITD_L to ITD_R at ISIs between approximately $\pm 50 \mu$ sec. Although Figures 6–8 show some asymmetries to the neural representations of these phantoms, in each case they are closely predicted by the correlation surface. Thus the neural responses seem to reflect accurately information contained in the ear-input signals. There is some indication, however, as was indicated in reference to Figure 5 above, that the signal-to-noise ratio of the neural response is enhanced over the binaural cross-correlation. Many cells showed an inhibition below their spontaneous firing levels, and often a rebound after stimulus offset, at ISIs and speaker-array locations that resulted in low binaural cross-correlations (*black areas* to either side of the central phantom in Figs. 6–8).

Each spatial response surface discussed above shows the expected pattern of activity in response to identical noise bursts. Figure 3 predicts that the responses to uncorrelated noise bursts should show no phantom sources and little or no modulation of the responses as the ISI is changed. Figure 9 shows the response of a cell to uncorrelated sounds. Its response to correlated sounds is shown in Figure 8. The spatial correlation surface looks quite similar to that predicted by Figure 3 and even more so to the surface of values extracted for $\tau = -20 \mu$ sec with uncorrelated noises (not shown). It should be noted, however, that in both the recorded neural response and the correlation data, the peak for

Figure 5. Responses of a single space-specific neuron. Two speakers, separated in azimuth by 55° , emitted identical noise bursts with various interstimulus delays (ISIs). Three differing speaker configurations are represented in *A–C* and are shown to the left of each set of plots (*RF* and *arrow* indicate the center of the receptive field of the cell at -5° azimuth). The left plots present the times of occurrence of action potentials for each of five repetitions of the stimulus pair with each interstimulus interval tested (enclosed by brackets on the abscissa). The time of stimulus presentation runs from bottom to top on the ordinate axis. These data are replotted on the right as spike rates normalized to the maximum rate recorded at any speaker configuration (filled circles, left ordinate scale). On these same axes, we plot the binaural cross-correlation at $\tau = 0$ (open circles, right ordinate scale) for each of the ISIs. These values were calculated from microphone recordings as in Figure 2, but with the speakers located as shown in the schematic drawings to the left.

the more centrally located speaker (the right speaker for this cell) was noticeably stronger than for the more peripheral speaker. This is also the case for correlation surfaces derived from the presentation of a single speaker. More centrally located speakers elicit stronger binaural cross-correlations. Thus, because of physical cues alone, spatial tuning curves measured in the free-field may seem more sharply tuned for neurons with more centrally located receptive fields (Knudsen and Konishi, 1978a).

We return for a moment to responses to identical sounds. Unlike the representations in Figure 2, neither the correlation surfaces nor the neural response surfaces show evidence of phantom sources located peripheral to the two real sources. Most of our cells had best azimuths within the frontal 45° or 50° of space. To test the responses of such cells to phantoms located outside the

two speakers, the array must be located well to one side or the other. At these array locations, the binaural cross-correlations are quite weak and thus it may be difficult to generate phantoms. In Figure 10, however, we show responses of a neuron with a broad receptive field, centered at -70° of azimuth. By placing the speaker array at locations near the owl's midline, we can present relatively strong phantoms that at some ISIs appear peripheral to the two speakers and fall within the receptive field of the cell. This demonstrates that phantom sources occurring peripheral to the two speaker locations can indeed be imaged on the space map.

DISCUSSION

Previously we described the response of space-specific neurons under reverberant conditions in which the ISI between the direct

Figure 6. *A*, Spatial correlation surface obtained with identical sounds at speaker-array azimuths from -90 to $+90^\circ$, inclusive, in steps of 5° . For each ISI and at each array location the weighted sum of values of the binaural cross-correlation was plotted for $\tau = -30, -20, -10, 0$, and $+10$ μsec (weights: 0.33, 0.66, 1.0, 0.66, 0.33). The figures are plotted to simulate the speaker array centered at 0° azimuth (speakers at $\pm 27.5^\circ$, indicated by *dashed lines*). *B*, Spatial response surface for the same space-specific neuron, the response of which is shown in Figure 5, and covering the same range of ISIs and speaker locations as shown in *A*. Notice the strong similarity in overall patterns between the correlation surface in *A* and the neural response in *B*. *C*, Single-speaker receptive field, centered at approximately -5° (approximately -10 μsec ITD).

sound and echo were considerably longer (0.5 and 5.0 msec) than those studied presently. Binaural cross-correlations predict that at these delays, the two sources would be represented with equal strengths. The neuronal response to the lagging sound, however, was found to be suppressed, leaving a stronger image of the leading source on the map (Keller and Takahashi, 1996). Thus, it

seems that when the ISI is long, the binaural signals contain the images of two sources but a neural mechanism reduces the image of the later source. Lateral inhibition, which has been reported in the owl's space map (Knudsen and Konishi, 1978b; Fujita and Konishi, 1991), may play a role in the suppression. The present results show that two sources are imaged with equal strength when the ISI is <0.5 msec, suggesting that the inhibition seen at the long ISIs is inoperative. At the shortest ISIs, near 0 msec, the cells respond as if there were a single phantom source located midway

Figure 7. Spatial correlation surface (*A*) and spatial response surface (*B*) for a space-specific neuron whose receptive field is centered at -25° azimuth (approximately -57.5 μsec ITD; *C*). Presentation the same as in Figure 6, with a slightly smaller range of azimuths tested. The binaural cross-correlation for each ISI and speaker-array location was calculated as the weighted sum of values for $\tau = -90, -80, -70, -60, -50, -40$, and -30 (weights: 0.25, 0.5, 0.75, 1.0, 0.75, 0.5, 0.25). The figures are plotted to simulate the speakers located at $\pm 22.5^\circ$ azimuth.

Figure 8. Spatial correlation (*A*) and spatial response (*B*) surfaces for a space-specific neuron whose receptive field was centered at -10° (approximately $-23 \mu\text{sec}$ ITD; *C*). Presentation the same as in Figure 6, with a smaller range of measured azimuths. The cross-correlation for each ISI and speaker-array location was calculated as the weighted sum of values for $\tau = -40, -30, -20, -10, 0$ (weights: 0.33, 0.66, 1.0, 0.66, 0.33). The figures are plotted to simulate two speakers located at $\pm 27.5^\circ$ azimuth.

between the two real sources. Two distinct phantoms can theoretically be distinguished only if the ISI exceeds the half-width of the spatial tuning curve of the cell (in microseconds of ITD).

Comparisons with behavior

To what extent does the binaural correlation surface represent the owl's perceptual experience? When owls are presented with two speakers emitting identical noises activated with a 1–10 msec delay, the owls make a rapid saccadic head-turn to face the leading speaker, suggesting that they localize but a single source (Keller and Takahashi, 1996). This behavior is reminiscent of the

precedence effect in humans (Wallach et al., 1949; Haas, 1951; Blauert, 1983). When the speakers are activated simultaneously, the owls turn their gaze upward from a reference speaker at foot-level to look at the space between the two speakers, suggesting that they perceive a single centrally located target (Keller and Takahashi, 1996). This too is consistent with human psychophysical data (Blauert, 1983) and with the neuronal responses described above (Figs. 6–8; Takahashi and Keller, 1994). Because delays between 0 and 1 msec were not tested in the earlier behavioral study, we cannot address the owl's perception at these short delays. However, given the complexity of the neural representation of reverberant environments in the ICx, and the evidence that the space map is necessary for spatial hearing in the owl (Wagner, 1993), it is clear that the behavior of the owl is also bound to be complex.

Data from human listeners under reverberant conditions are extensive (for review, see Blauert, 1983), and it is informative to consider their responses, despite the obvious differences in the morphology of owls and humans. Summing localization takes effect in humans when the interval between the direct sound and echo is less than approximately 1 msec. Summing localization is generally believed to be attributable to the superposition of the direct sound and echo and does not occur when the sounds of the two speakers are uncorrelated (Damaske, 1969/70). Typically, human subjects perceive a single sound source at a position located between the two speakers but closer to the leading source. If the two sources are activated with no delay, human listeners perceive a single source midway between the two actual sources. Although most subjects report a single source, careful listeners have reported multiple targets and have perceived targets located beyond the speakers themselves (Blauert, 1983). The schematic representation of Figure 2*B* shows the presence of strong peaks in binaural cross-correlation that extend diagonally across τ as the ISI is changed. One of the diagonal lines crosses from $\tau = \text{ITD}_l$ through $\tau = 0$ to $\tau = \text{ITD}_r$ as ISI is changed from small negative values to small positive values. This diagonal could

Figure 9. Spatial response surface obtained with uncorrelated sounds for the same cell whose response is shown in Figure 8. Same weights and speaker locations as in Figure 8. Note the lack of crossing diagonals as was seen in response to phantom sources when identical sounds were broadcast from the two speakers. The more peripheral (*left*) speaker is represented more weakly than is the more central (*right*) speaker.

Figure 10. Phantom sources exist at azimuths peripheral to the two loudspeakers at certain ISIs. *A*, Spatial correlation surface for a cell whose broad receptive field was centered near -70° ($-161 \mu\text{sec}$ ITD; weights approximate the single-speaker curve in *C*). At ISIs between -180 and -100 , and $+40$ and $+110 \mu\text{sec}$, two areas of high correlation occur between approximately -55° and -80° (~ 125 – $185 \mu\text{sec}$ ITD). The figures are plotted to simulate two speakers located at $\pm 27.5^\circ$ azimuth. *B*, Spatial response surface for the same cell. Note broader ISI scale. *C*, Single-speaker spatial tuning curve.

represent the binaural cues that allow perception of the commonly reported phantom target, which migrates from one side to the other as ISI is varied. It is also clear, however, that the complementary diagonal, which has the opposite trajectory, as well as the real sources are available for localization at these near-zero delays. Furthermore, the diagonals extend beyond the loci of the real targets. Perhaps these regions of high binaural

correlation account for the multiplicity of targets and for their extreme perceived loci that are reported by the careful listeners.

Nevertheless, the most common experience is a single target. It is likely that the difference between the acoustical and neurophysiological images and the common perception is attributable to the involvement of higher perceptual and cognitive centers that generate the ultimate perception of the auditory scene or pick the targets to which attention shall be directed. A neural image derived from a cross-correlation-like mechanism, such as that displayed in the owl's ICx, can serve as the source of spatial information for these higher processes.

Comparisons with other neurophysiological studies

The response of auditory neurons to simulated reverberant conditions has been studied in a number of nuclei and in various species (cat: Whitfield et al., 1972; Cranford and Oberholtzer, 1976; Yin, 1994; rabbit: Fitzpatrick et al., 1995; rat: Kelly, 1974; mouse: Wickesberg and Oertel, 1990; bat: Yang and Pollak, 1994, 1995; cricket: Wytenbach and Hoy, 1993; barn owl: Keller and Takahashi, 1996). Most of these studies have examined the effects of delays on the order of milliseconds, which are much longer than those used in the present study. Generally, the studies have reported that the response of the neuron to a sound in its receptive field can be suppressed by an earlier sound, and the authors have drawn analogies with the phenomenon of the precedence effect.

Only the study of Yin (1994), in the IC of the cat, has examined explicitly the neural basis of summing localization. The IC neurons of the cat, like those of the owl, have spatial receptive fields based on their sensitivity to binaural cues. Yin presented two stimuli in rapid succession ($< 2 \text{ msec}$ ISI) from either two free-field speakers or dichotically with two different ITDs. In several cells, a plot of the response of the cell as a function of ISI was quite similar to the profile of the receptive field of the cell along the azimuth, or, for dichotic stimuli, to its ITD-sensitivity function. As Yin (1994) points out, this would be expected if an auditory image moved across the receptive field as the ISI was varied, just as described in human summing localization. Furthermore, Yin found that increasing the sound level of the lagging click would shift the ISI response curve, as though a source was now closer to the louder, lagging sound. This time-intensity trade is also seen in human summing localization (Snow, 1954).

Our results thus are qualitatively consistent with those of Yin (1994). The graph at the upper right of Figure 5 (*solid dots*), which plots firing rate as a function of ISI when the two speakers are placed almost symmetrically about the receptive field, shows that the resulting function is similar to the single-speaker spatial response function of the cell (Fig. 6C). By systematically changing the ISI, the peak of the correlation corresponding to the phantom source travels along a diagonal, schematically shown in Figure 2B, traversing the receptive field of the neuron (mapped along the τ -axis in Fig. 2B) at a rate of $1 \mu\text{sec}$ ITD for each microsecond of change of ISI. The range of ISIs over which the phantom falls within the receptive field of a cell is typically smaller for the owl than for the cat (Yin, 1994) and will depend on the receptive-field width (in microseconds of ITD) as well as the spread of the phantom along the ISI axis. This difference between the results of Yin (1994) and our own is expected because receptive field widths expressed as microseconds of ITD are typically narrower in owls than in cats (Moiseff and Konishi, 1981; Yin and Chan, 1988), attributable to the owl's ability to phase-lock at higher frequencies (Sullivan and Konishi, 1984). Also, the spread of the phantom

image can be affected by numerous factors, but perhaps most importantly in the present instance, by our use of smaller inter-speaker distances resulting in higher binaural correlations.

Yin (1994) does not propose a specific mechanism for the behavior of the cells in the cat IC during simulated summing localization. He draws an analogy to backward masking, however, pointing out that in summing localization, as in backward masking, a later sound can influence the perception of an earlier sound source. An earlier study (Carney and Yin, 1989) showed inhibition of IC responses to monaural clicks that was consistent with a backward masking effect and might reasonably explain results shown by Yin (1994). Our results suggest that in the owl at least, the similarity between the ISI and single-speaker functions depends on the same use of a binaural cross-correlation-like mechanism whether there is one source or more than one. This idea may be extended to the time-intensity trade in which the perceived locus of a target in summing localization can be biased toward the louder source (Snow, 1954). In an earlier study, we demonstrated that when the sounds of two speakers are produced simultaneously and are identical except for overall amplitude, the neural image on the space map is biased toward the louder speaker (Takahashi and Keller, 1994). This result too is predicted from the superposition of the waveforms of the two sources in the ears and the binaural cues computed from these signals (Bauer, 1961; Blauert, 1983; Takahashi and Keller, 1994). The selective imaging of, or attention to, only one of several possible sources may involve an inhibition of later responses similar to that underlying a precedence-like effect.

REFERENCES

- Albeck Y, Konishi M (1995) Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals. *J Neurophysiol* 74:1689–1700.
- Bauer BB (1961) Phasor analysis of some stereophonic phenomena. *J Acoust Soc Am* 33:1536–1539.
- Blauert J (1983) Spatial hearing. Cambridge, MA: MIT.
- Carney LH, Yin TCT (1989) Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: excitatory and inhibitory components. *J Neurophysiol* 62:144–161.
- Carr CE, Konishi M (1990) A circuit for detection of interaural time differences in the brainstem of the barn owl. *J Neurosci* 10:3227–3246.
- Cranford JL, Oberholtzer M (1976) Role of neocortex in binaural hearing in the cat. II The “precedence effect” in sound localization. *Brain Res* 111:225–239.
- Damaske P (1969/70) Richtungsabhängigkeit von Spektrum und Korrelationsfunktionen der an den Ohren empfangenen Signale. *Acustica* 22:191–204.
- Fitzpatrick DC, Kuwada S, Batra R, Trahiotis C (1995) Neural responses to simple simulated echoes in the auditory brain stem of the unanesthetized rabbit. *J Neurophysiol* 74:2469–2486.
- 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.
- Haas H (1951) Über den einfluss eines Einfachechos auf die Hörsamkeit von Sprache. *Acustica* 1:49–58.
- Jeffress LA (1948) A place theory of sound localization. *J Comp Physiol Psychol* 41:35–39.
- Keller CH, Takahashi TT (1996) A precedence effect in the owl's auditory space map? *J Comp Physiol [A]* 178:499–512.
- Kelly JB (1974) Localization of paired sound sources in the rat: small time differences. *J Acoust Soc Am* 55:1277–1284.
- Knudsen EI, Konishi M (1978a) A neural map of auditory space in the owl. *Science* 200:795–797.
- 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 [A]* 133:13–21.
- Kuwada S, Yin TCT (1983) Binaural interaction in low-frequency neurons in inferior colliculus of the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function. *J Neurophysiol* 50:981–999.
- Moiseff A (1989) Binaural disparity cues available to the barn owl for sound localization. *J Comp Physiol [A]* 164:629–636.
- Moiseff A, Konishi M (1981) Neuronal and behavioral sensitivity to binaural time differences in the owl. *J Neurosci* 1:40–48.
- 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 B McA, Cherry EC (1957) Mechanism of binaural fusion in the hearing of speech. *J Acoust Soc Am* 29:973–987.
- Snow W (1954) The effects of arrival time on stereophonic localization. *J Acoust Soc Am* 26:1071–1074.
- 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, Keller CH (1994) Representation of multiple sound sources in the owl's auditory space map. *J Neurosci* 14:4780–4793.
- Wagner H (1993) Sound-localization deficits induced by lesions in the barn owl's auditory space map. *J Neurosci* 13:371–386.
- Wagner H, Takahashi TT, Konishi M (1987) Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *J Neurosci* 7:3105–3116.
- Wallach H, Newman EB, Rosenzweig MR (1949) The precedence effect in sound localization. *Am J Psychol* 52:315–336.
- Whitfield IC, Cranford J, Ravizza R, Diamond IT (1972) Effects of unilateral ablation of auditory cortex in cat on complex sound localization. *J Neurophysiol* 35:718–731.
- Wickesberg RE, Oertel D (1990) Delayed, frequency-specific inhibition in the cochlear nuclei of mice: a mechanism for monaural echo suppression. *J Neurosci* 10:1762–1768.
- Wytenbach RA, Hoy RR (1993) Demonstration of the precedence effect in an insect. *J Acoust Soc Am* 94:777–784.
- Yang L, Pollak GD (1994) The roles of GABAergic and glycinergic inhibition on binaural processing in the dorsal nucleus of the lateral lemniscus of the mustache bat. *J Neurophysiol* 71:1999–2013.
- Yang L, Pollak GD (1995) Binaural inhibition in the dorsal nucleus of the lateral lemniscus of the mustache bat affects responses for multiple sounds. *Auditory Neurosci* 1:1–17.
- Yin TCT (1994) Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. *J Neurosci* 14:5170–5186.
- Yin TCT, Chan JCK (1988) Neural mechanisms underlying interaural time sensitivity to tones and noise. In: *Auditory function* (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 (1984) Neuronal mechanisms of binaural interactions. In: *Dynamic aspects of neocortical function* (Edelman GM, Gall WE, Cowan WM, eds), pp 263–313. New York: Wiley.
- 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.