Multisensory Enhancement in the Optic Tectum of the Barn Owl: Spike Count and Spike Timing

Yael Zahar, Amit Reches, and Yoram Gutfreund
Department of Physiology and Biophysics, The Rappaport Faculty of Medicine and Research Institute, Technion, Haifa, Israel

Submitted 9 November 2008; accepted in final form 2 March 2009

Zahar Y, Reches A, Gutfreund Y. Multisensory enhancement in the optic tectum of the barn owl: spike count and spike timing. J Neurophysiol 101: 2380–2394, 2009. First published March 4, 2009; doi:10.1152/jn.91193.2008. Temporal and spatial correlations between auditory and visual stimuli facilitate the perception of unitary events and improve behavioral responses. Here we studied responses of multisensory neurons in the barn owl’s optic tectum (the avian homologue of the superior colliculus) to visual, auditory, and bimodal stimuli. We specifically focused on responses to sequences of repeated stimuli. First, we report that bimodal stimulation tends to elicit more spikes than in the responses to its unimodal components (a phenomenon known as multisensory enhancement). However, this tendency was found to be history-dependent; multisensory enhancement was mostly apparent in the first stimulus of the sequence and to a much lesser extent in the subsequent stimuli. Next, a vector-strength analysis was applied to quantify the phase locking of the responses to the stimuli. We report that in a substantial number of multisensory neurons responses to sequences of bimodal stimuli elicited spike trains that were better phase locked to the stimulus than spike trains elicited by stimulating with the unimodal counterparts (visual or auditory). We conclude that multisensory enhancement can be manifested in better phase locking to the stimulus as well as in more spikes.

INTRODUCTION

Many events in everyday life are registered in parallel by the visual and auditory senses. Nevertheless, our perception of the external world is unitary and coherent. This raises a challenging question known as the binding problem: how are these distributed representations integrated in the brain to form one homogeneous percept? Binding has been suggested to depend on several factors; among them are spatial agreement, temporal correlations, and cognitive processes (Spence 2007). In particular, temporal and spatial correlations between auditory and visual stimuli facilitate the perception of unitary events and improve behavioral responses (Frassinetti et al. 2002; Jiang et al. 2002; Narins et al. 2005; Whitchurch and Takahashi 2006).

Several aspects of visual-auditory binding were shown to be correlated with responses of multisensory neurons in the superior colliculus (SC), a structure in the midbrain involved in attentive and orientation behaviors. Neural responses to spatiotemporally coinciding multisensory stimuli exhibit stronger responses (Meredith and Stein 1986; Wallace et al. 1996) and reduced latencies (Rowland et al. 2007) than their unimodal counterparts. Here we study visual-auditory integration in multisensory neurons of the barn owl’s optic tectum (OT; the avian homologue of the SC).

The barn owl is a nocturnal animal that has evolved highly precise visual and auditory systems (Harmening et al. 2007; Knudsen et al. 1979) to detect small prey in acoustically noisy and dimly lit environments. Neurons that respond to both visual and auditory inputs are more abundant in the OT of the barn owl than in the SC of the cat (an animal model commonly used to study visual-auditory integration), existing in the superficial as well as in the deep layers (Knudsen 1982). The tuning of neurons for auditory space in the barn owl is sharper than in any other species studied. As a result the alignment between visual and auditory receptive fields in multisensory neurons is exceptionally accurate (Knudsen and Brainard 1991). The OT of the barn owl thus provides us with an outstanding model system for studying cross-modal integration at the single-cell level.

Most studies on visual-auditory interactions at the single-neuron level use simple single stimuli. Here we study responses of multimodal neurons to long sequences with varying interstimulus intervals (ISIs). We show a tendency for bimodal enhancement as been defined by Stein and Meredith (1993). However, this enhancement was common only for the responses to the first event in a sequence. Subsequent events rarely demonstrated cross-modal enhancement even for stimulus rates as low as 1.25 Hz. To assess the possibility that, at subsequent events, the cross-modal interaction is manifested in spike timing rather than spike counts, we quantified phase locking to the stimulus using vector-strength analysis. We show that in a large portion of the neurons phase locking to a coherent bimodal stimulus was significantly better than the phase locking to the unimodal stimuli or bimodal but incongruent stimulus.

METHODS

Eight barn owls (Tyto alba) were studied in this research. All owls were hatched in captivity and were raised and kept in a large flight cage equipped with perching spots and nesting boxes. The owls were provided for in accordance with the guidelines of the Technion Institutional Animal Care and Use Committee.

Surgical procedures

Owls were prepared for repeated electrophysiological experiments in a single surgical procedure: the owl was anesthetized with halothane or isoflurane (2%) and nitrous oxide in oxygen (4:5). The scalp was prepared with Betadine povidone-iodine (10%) surgical scrub. An incision was made in the scalp, and the skull was scrapped clean in two places. At one place, just anterior to the neck muscle, a small stainless steel plate with a protruding bolt was cemented to the skull using dental cement. At a second place, determined by stereotaxic...
coordinates, a craniotomy was performed, and a threaded recording chamber was cemented to the skull. The wound was sutured, and incisions were infused with lidocaine. After surgery the animal recovered overnight in an individual cage and then released back to its home cage.

Electrophysiological recordings

Before each electrophysiological session the owl was moved to an individual cage and deprived of food overnight. At the beginning of the recording session, the owl was anesthetized briefly with halothane or isoflurane (1.5%) and nitrous oxide in oxygen (4:5), wrapped in a soft leather jacket, and placed in a stereotaxic apparatus at the center of a double wall sound-attenuating chamber (internal size: 2.05 × 1.7 × 1.95 m; IAC, Bronx, NY) lined with echo suppressing foam. The head was bolted to the stereotaxic apparatus and aligned using retinal landmarks (as described in Gold and Knudsen 2000). Once the bird was secured in place, halothane was removed from the gas mixture, and the bird was maintained on a fixed mixture of nitrous oxide and oxygen (4:5). The head chamber was cemented to the skull. The wound was sutured, and we did not immobilize or control for eye movements. Electrophysiological recordings began ≥1 h after removal of halothane or isoflurane, allowing enough time to recover from the anesthetic agent. A Tucker-Davis Technologies (Alachua, FL) System3 and an on-line spike sorter (MSD, Alpha-Omega) were used to record and isolate action potentials from single neurons or a small cluster of neurons (multiunit recording). Multiunit recordings were obtained by manually setting a threshold consistently selecting the largest unit waveforms in the recorded site. Single units were isolated using a template-based sorting. The spike sorter presents a histogram of the squared errors between the template and the detected spike. We required the histogram to have a sharp well-distinguished peak, signifying the presence of a homogeneous group of spike shapes similar to the template. We also verified that the ISI histogram showed an absolute refractory period of ≥1 ms. Because of the characteristic bursty firing of many tectal sites, adequate template matching was difficult to obtain. Therefore most of the recordings in our data set were multiunit recordings. To verify that the main conclusions of the paper are not an outcome of multiunit recordings, we analyzed separately data from well isolated single units. We concluded that there was no qualitative difference between the two signals. Results from single-unit recordings are shown in Figs. 5 and 8, C and D. At the end of each recording session the chamber was treated with chloramphenicol ointment (5%) and closed. Theowl was then returned to its home flying cage.

Targeting of the OT

The identification of the OT was based on stereotoxic coordinates and expected physiological properties. The OT was recognized by characteristic bursting activity and spatially restricted visual receptive fields (Knudsen 1982). Position within the OT was determined based on the location of the visual RF.

Stimulus presentation

Auditory stimuli were produced using custom Matlab program, converted to analog (48-kHz sampling rate, RP2; Tucker-Davis Technologies) and transduced by a pair of matched miniature earphones (ED-1914; Knowles, Itasca, IL). The earphones were placed in the center of the ear canal ~8 mm from the tympanic membrane. The amplitude and phase spectra of the earphones were equalized within ±2 dB and ±2 μs between 2 and 12 kHz by computer adjustment of the stimulus waveform. Sound levels were controlled by two independent attenuators (PA5; Tucker-Davis Technologies), and are reported as average binaural sound intensity (ABSI) relative to a fixed sound pressure level. The acoustic stimuli consisted of broadband noise (3–12 kHz; finite impulse response filter, order = 70).

Visual stimuli were computed in Matlab, using the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997) and projected (refresh rate 72 Hz, XD400U; Mitsubishi) on a screen inside the sound-attenuating chamber (screen size 150 × 115 cm, 1.5 m away from the owl). The projector was positioned outside the chamber, projecting the image through a double-glass window.

In each recording site, the center of the visual receptive field was first mapped by moving a visual stimulus over the screen (computer generated or a spot of light projected manually on the screen). The area of best response was identified based on listening to the discharge rate of the responses. Later the tuning curves for interaural time difference (ITD), interaural level difference (ILD), and ABSI were recorded by varying a single parameter (ITD, ILD, or ABSI) while holding all other parameters constant. The value of the tested parameter was varied randomly in stimulus sets that were repeated 10–20 times. Unit responses to an acoustic stimulus were quantified as the number of spikes in a given time window after stimulus onset minus the number of spikes during the same amount of time immediately before stimulus onset (baseline activity). The width of tuning curves was defined as the range over which responses were >50% of the maximal response; best ITD or ILD was the mid-point of this range. Response threshold was determined as the lowest ABSI that elicited a significant response.

Stimulus design

To test bimodal integration, 13 different 8-s-long sequences of visual, auditory, or a combination of auditory and visual stimuli (bimodal stimuli) were randomly interleaved with gaps of 2-s in-between (see Fig. 1B). Each stimulus block was repeated 10–15 times. The visual stimuli consisted of a dark circle on a brighter background positioned at the center of the visual receptive field. The size of the circle was 1.9°, luminance of background screen was 17 cd/m², and luminance of circle was 8–15 cd/m², set to be slightly above response threshold as estimated by listening to the spiking activity of the recorded site. The stimulus appeared on the screen at onset, moved a short distance of 1.9° for a duration of 166 ms and disappeared (Fig. 1A). This event was presented at six different repetition rates varying between 0.5 and 4.25 Hz (Fig. 1B, table). The auditory stimulus consisted of a sequence of broadband sounds (2–12 kHz) with a fixed half-sine wave envelope, 166 ms wide (Fig. 1A). The ITD and ILD values were set to the best ITD and best ILD of the recorded site. The ABSI level was chosen as 10–20 dB above unit threshold.

We presented two tests, one in which the auditory stimulus was presented at a slow rate (1.25 Hz) and the other at a higher rate (2.75 Hz). In the bimodal condition, the visual sequences (0.5–4.25 Hz) were combined with the auditory sequence (1.25 or 2.75 Hz). The visual and auditory sequences were initiated at the same time. In one case, the repetition rates matched and thus the sequence was bimodally synchronized (see Fig. 1B, table). The matched bimodal combinations are termed the congruent tests (1.25 or 2.75 Hz); the bimodal combinations of mismatched frequencies are termed incongruent sequences. The frequencies of 2.75 and 1.25 Hz were chosen to avoid harmonics matching between the auditory and the congruent visual frequencies (0.5, 1.25, 2, 2.75, 3.5, and 4.25 Hz). A photodetector on the screen was used to measure the actual synchronicity between visual and auditory sequences. In our setup the visual stimulus appeared on the screen within ±14 ms from the onset of the auditory stimulus. This synchrony error was ignored in the present work.

Data analysis

To quantify the degree of synchronization of the spike train with the fundamental frequency of the amplitude-modulated stimulus, vector-
the two unimodal responses (VSA and VSV): enhancement index of the assessed frequency, Takahashi 2005) as follows

\[
\text{VS} = \frac{1}{n} \sqrt{x^2 + y^2}; x = \sum_{i=1}^{n} \cos \theta_i; y = \sum_{i=1}^{n} \sin \theta_i; \theta_i = 2\pi \frac{t_i}{T} \quad (1)
\]

Where \(t_i\) is the post stimulus time of spike \(i\), \(T\) is the cycle period of the assessed frequency, \(\theta_i\) is the phase of the spike \(i\) in the cycle period \(T\), and \(n\) is the total number of spikes in the spike train. The VS may vary from 0 to 1 with larger values implying strong phase locking of the spike train to the period \(T\), and values close to 0 indicating no synchronization. For each stimulus type, VSs were calculated for frequencies ranging between 1 and 10 Hz, including the frequency of the stimulus. Rayleigh statistic (Mardia and Jupp 2000) was used to assess whether the VS at the frequency of the stimulus was significantly larger than zero indicating that the spike train is modulated in-phase with the modulation of the stimulus. To test whether the VS of the response to the matched bimodal stimulus was significantly different from the VSs of the unimodal responses, we used the bootstrap method (Mardia and Jupp 2000). From the data of each test (10–15 repetitions), we drew 2,000 bootstrap data samples of visual, auditory, and bimodal responses and calculated the distribution of the bimodal VS minus the VS of the maximal unimodal response. We tested the null hypothesis that the difference between the VS of the bimodal response and the VS of the unimodal response is zero.

An enhancement index was calculated to compare the VS of the bimodal response (VS\(_{AV}\)) with the maximal vector strength between the two unimodal responses (VS\(_A\) and VS\(_V\)): enhancement index = VS\(_{AV}\)−max(VS\(_A\), VS\(_V\)). In the congruent tests, the VSs were calculated for the frequency of the stimulus (1.25 or 2.75 Hz). In the incongruent tests, each unimodal VS was calculated for the frequency of its own stimulus. The incongruent bimodal VS was calculated for the frequency of the unimodal sequence that gave rise to the largest VS.

In addition we calculated the spike count (SpC) enhancement index, introduced by Meredith and Stein (1983), defined as \(\text{SpC}_{AV^*} = \text{max}(\text{SpC}_A, \text{SpC}_V)/\text{max}(\text{SpC}_A, \text{SpC}_V)\) where \(\text{SpC}_{AV}\) is the spike count of the bimodal response, \(\text{SpC}_A\) is the spike count of the auditory response, and \(\text{SpC}_V\) is the spike count of the visual response. Spike counts are defined as the number of spikes counted in a specific time window relative to the stimulus minus the number of spontaneous spikes measured in the equivalent time window prior to the stimulus. To avoid erroneously large indices as a result of division by close to zero values, we removed from the distributions in Fig. 3 and from the statistical analysis the several cases where the maximal unimodal response was <0.2 spike/s (7 cases in Fig. 3A; 14 cases in Fig. 3B; 3 cases in Fig. 3C; 10 cases in Fig. 3D; 6 cases in Fig. 3E; 5 cases in Fig. 3F).

To compare the bimodal responses with that expected by simple summation of the unimodal responses, the visual and auditory unimodal spike trains were merged and transformed into bin-representation (bin size = 20 ms). The spontaneous activity (the average number of spikes per bin in the prestimulus period) was removed from each bin to avoid summation of background activity. The VS of the histogram of the sum was then calculated and compared with the VS of the bimodal response.

A bootstrap technique was applied to assess the significance of the difference between the bimodal response and the sum of the unimodal responses (Fig. 9). In each trial, the number of spikes at a particular time window in the bimodal, visual, or auditory response was measured. From these distributions, 1,000 data sets were generated (resampling with replacements). For each data set, the average bimodal

![Fig. 1. Visual and auditory stimuli used in this study. A, left: an example of the visual display. Right: examples of the waveform of the auditory stimulus (bottom) and the time course of the visual stimulus (top). In this example, the visual and auditory stimuli were synchronized at 1.25 Hz. B: time line of the experimental test. In each test, stimuli were randomly presented from 13 different stimulus combinations described in the tables below. Each sequence presentation was 8-s long followed by a 2-s silent gap. The gray boxes in the tables designate the congruent combinations.](http://jn.physiology.org/Download/10.1152/jn.00027.2009)}
response was reduced from the sum of the average unimodal responses. The null hypothesis that the mean of the differences is zero was tested.

A bootstrap technique was also applied to assess the population difference between the bimodal response and the sum of the unimodal responses (Fig. 10, E and F): the population of 175 recording sites was resampled with replacements to obtain one thousand datasets. For each data set, the time to peak and the duration of the average bimodal response (defined as the duration of response >50% of maximal response) was measured. Similarly the time to peak and duration of the sum of the unimodal responses were measured. t-tests were used to test the null hypotheses that the differences between the parameters (time to peak and response duration) were zero.

**Simulating spike trains**

A Poisson spike generator was used to simulate visual, auditory, and bimodal spike trains. First the responses to all the events in the sequence were collapsed into one post stimulus time histogram (bin size = 2 ms, duration of histogram = stimulus cycle). Then all recording sites were averaged. This was repeated for visual, auditory and bimodal stimuli giving the estimated firing rates (FR(t)). For each time bin in the histogram, a random number was generated between 0 and 1 [n(t)]. If FR(t)*2 ms > n(t), a spike was generated at the time bin t. This procedure was repeated for all time bins to create a simulated response.

**RESULTS**

Responses of 265 recording sites in the OT were sampled. Recording sites spanned both bursty layers [superficial and intermediate layers (Knudsen 1982) and nonbursty layers deep layers (Knudsen 1982)]. All recording sites were from the anterior part of the OT having visual receptive fields between left and right 20° and up and down 20° relative to the center of the visual field. Recording sites were classified as multisensory if the VS analysis revealed a significant phase locking to both the visual and the auditory stimuli presented alone (Rayleigh test, P < 0.05) or if the VS of the bimodal stimulus was significantly different from the VS of the most effective single modality (2-tailed bootstrap test, P < 0.05). Based on this criterion 175 of 265 sites were multisensory, the remaining sites were not included in the following analyses.

**Spike count analysis**

As a first step to analyze the neural responses, we calculated the SpC enhancement index, estimating the difference between the bimodal response and the responses to the unimodal components of the stimulus. Examples from two recording sites are displayed in Fig. 2, A and B, both showing responses to the congruent 1.25-Hz test. In these examples, the enhancement index calculated throughout the time of the stimulus (8 s) was positive (SpC enhancement index equals 2.18 in the upper example and 0.77 in the lower example). However, as can be seen in the raster plots and in the corresponding average response histograms (Fig. 2, C–F), the bimodal enhancement was apparent in the response to the first event in the sequence and not in subsequent responses. Therefore in the following population analysis, we analyzed the responses in two separate time windows: in a time window of 0–800 ms containing only the first event and in a time window of 0.8–8 s containing the rest of the events.

Figure 3, A and B, shows the distributions of the enhancement indices measured from responses to the congruent 1.25-Hz test. The distribution of the SpC enhancement indices for the first event of the sequence (Fig. 3A) was significantly shifted toward positive...

---

**FIG. 2.** Spike count bimodal enhancement. A: a raster plot showing an example of 10 responses of a single site to the 1st 5 visual stimuli in a sequence of 1.25 Hz (bottom), 10 responses to the congruent auditory sequence (middle), and 10 responses to the bimodal combination of the visual and auditory sequences (top). The black bars at the bottom designate the time course of the stimuli (each bar is 166 ms). The vertical line marks the time of stimulus onset. B: a raster plot showing an example from a different site of responses to the same stimuli as in A. Format as in A. C: average responses of the example (A), measured from the 1st 800 ms of the auditory, visual and bimodal stimuli. Error bars show the SE. D: average responses of the same example measured from the time window starting at the onset of the 2nd event (800–3,200 ms). Error bars show the SE. E and F: average responses of the site shown on the left (B). Format as in C and D.
the congruent sequences. In summary, despite the clear influence of the additional modality. A significant tendency to positive enhancement indices for the first event (0 – 0.36 s) of the congruent 2.75-Hz tests was evident in the distribution of SpC indices excluding the first event (Fig. 3D). Figure 3, E and F, shows the distributions of the enhancement indices calculated from results of the incongruent test in which the auditory sequence was 1.25 Hz and the visual sequence was 2.75 Hz. Indices were calculated from SpCs during the period of the 1st visual event (0–360 ms). Format as in B, inset: E: distribution of the SpC enhancement indices calculated from responses to the incongruent tests in which the auditory sequence was 1.25 Hz and the visual sequence was 2.75 Hz. Indices were calculated from SpCs during the period of 0.36–8 s. Inset: the enhancement indices in the 1st event plotted against the enhancement indices in the subsequent events. Format as in B, inset: F: distribution of the SpC enhancement indices, calculated from responses to the incongruent tests in which the auditory sequence was 1.25 Hz and the visual sequence was 1.25 Hz, is shown in Fig. 4. The distribution of the enhancement indices calculated from responses to the congruent 2.75-Hz stimuli without the period of the 1st event (time window of 0.36–8 s). Inset: the enhancement indices in the 1st event plotted against the enhancement indices in the subsequent events. Format as in B, inset: G: distribution of the SpC enhancement indices calculated from responses to the congruent 2.75-Hz stimuli without the period of the 1st event (time window of 0 – 800 ms). and visual responses, respectively. The distribution of the enhancement indices calculated from SpCs during the period of 0.36–8 s for the first event (Fig. 3D). Figure 3, E and F, shows the distributions of the enhancement indices calculated from results of the incongruent test in which the auditory sequence was presented at a rate of 1.25 Hz and the visual stimuli at a rate of 2.75 Hz. Note that the first stimulus in all tests was always congruent. The results of the analysis showed a similar tendency as observed in the congruent tests (1.25 Hz, Fig. 3, A and B, and 2.75 Hz, C and D). Counting the overall number of spikes without the first event did not reveal a clear influence of the additional modality. A significant tendency for bimodal enhancement (sign test; $P < 0.01$) in the first stimulus and a nonsignificant tendency (sign test; $P > 0.05$) in the remaining period were observed in all bimodal combinations tested (only 3 combinations are shown in Fig. 3). In summary, despite the clear tendency for bimodal enhancement during the first congruent event in the sequence, no population tendency for enhancement was observed during repetitive presentations of the stimuli even in the congruent sequences.

Interestingly, several recording sites showed significant suppression of the SpC in the first event ($t$-test, $P < 0.05$), but the number of significant suppressions was smaller than the number of significant enhancements (compare $\bullet$ on the left side of the histograms in Fig. 3 with $\square$ on the right side): 30 compared with 11 in the congruent 1.25-Hz test (Fig. 3A), 19 compared with 1 in the congruent 2.75 Hz test (C), and 16 compared with 7 in the incongruent case (E). In all three cases, the median of the sites with significant integration was larger than zero (sign test, $P < 0.01$). As can be seen in Fig. 3, B, D, and F, in several sites significant suppression was also registered during the later time window. However, unlike during the first time window, the number of sites significantly suppressed was not significantly smaller (sign test, $P > 0.05$) than the number of sites with significant enhancement.

The fact that most of the recordings were multiunit (see METHODS) raises the concern that the multisensory enhancement results might be governed by pooling together, in single recordings, independent unisensory units of different modalities. If this indeed was the case, we would expect the interaction to be additive (the sum of the unimodal responses should equal the bimodal response). To assess this, we estimated the difference of the bimodal response from the expected (additive) response by calculating the index: $|\text{AV} - (A + V)/(A + V)|$. AV, A, and V are defined as the SpCs of the bimodal, auditory, and visual responses, respectively. The distribution of the indices, calculated for a time window of 0–800 ms from the congruent 1.25-Hz tests, is shown in Fig. 4. The distribution clearly did not peak at zero (as would be the case for additive interaction): The peak of the distribution was negative (subadditive) with a tail stretching to the positive domain (superadditive).
To further address the issue of multi-versus single-unit recordings, we analyzed separately results from experiments in which clear single units were isolated \((n=34)\). We observed a similar tendency for SpC enhancement in the first event of the sequence (Fig. 5A; congruent 1.25-Hz test; sign test; \(P<0.01\)) and not in the subsequent events (0.8–8 s; congruent 1.25-Hz test, B and C). The median of the single-unit results of the first event was significantly larger (Mann-Whitney test at a significant level of 0.05 but not 0.01) from the median of the corresponding multiunit distributions (Fig. 5A compared with Fig. 3A), suggesting that the single-unit data better reflect the SpC enhancement.

**VS analysis**

An example from another recording site is shown in Fig. 6. In this example, the repetition rate of the visual, auditory and bimodal stimuli was 1.25 Hz. The firing rate fluctuated in accordance with the temporal modulation of the visual stimulus (Fig. 6A). However, responses from the same site failed to follow the temporal modulations of the auditory stimulus (Fig. 6B). Responses to the auditory stimulus were weak, displaying a phasic significant increase in firing rate only to the first stimulus of the sequence (\(t\)-test, \(P<0.05\)). During the subsequent period (0.8–8 s), a slight nonsignificant increase in the

**FIG. 6.** Example of a single tectal site response to a unimodal visual stimulus (A), to the matched unimodal auditory stimulus (B), and to the bimodal combination of the auditory and visual stimuli (C). The bottom part of each panel shows the post stimulus time histogram and the upper part displays the corresponding dot raster (10 responses to the same stimulus). The bars in A designate the time course of the stimuli (a sequence of 166-ms long events at a rate of 1.25 Hz). The SpC values shown in the box of each panel designate the average number of spikes per second in the poststimulus period (0–8 s) minus the average number of spikes per second in the prestimulus period.

**FIG. 5.** Summary of results from a subpopulation of well-isolated single units. A: distribution of the SpC enhancement indices calculated from responses to the first event in the congruent 1.25-Hz sequence (time window of 0–800 ms). B: distribution of the SpC enhancement indices calculated without the period of the first event (time window of 0.8–8 s). (A and B), neurons with significant enhancement or suppression of bimodal responses (\(t\)-test, \(P<0.05\)). *, distributions that were significantly shifted to positive values (sign test; \(P<0.01\)). C: for each neuron, the EI in the 1st event is plotted against the EI in the subsequent events. —, the points of equality; numbers in the top right corner show the number of neurons above and below the line.
overall firing rate was recorded that was not in synchrony with the temporal structure of the stimulus. Despite this shortfall to convey the temporal structure of the auditory sequence, the response to the matched bimodal sequence (Fig. 6C) was better phase locked to the stimulus than the response to the visual only sequence, displaying higher and sharper response peaks (compare Fig. 6, A with C). Interestingly, the average number of spikes elicited by the bimodal stimulus throughout the stimulation period (0–8 s) was not different from in the response to the visual stimulus (6.6 spike/s in the bimodal response vs. 6.3 spike/s in the visual only response). In this example, therefore, the interaction between visual and auditory inputs was manifested in spike timing (better phase locking) rather than in spike numbers.

To quantify the locking of the response to the temporal modulation of the stimulus we used the VS analysis (see METHODS). Figure 7 shows the results of this analysis from a single recording site (different from the site shown in Fig. 6). Each plot depicts results from 1 of the 13 stimulus combinations (see Fig. 1B). The auditory repetition rate in this example was 1.25 Hz. In the responses to the low visual repetition rates (0.5–2.75 Hz; Fig. 7, top) and in the response to the auditory only stimulus (bottom), the modulation frequency of the stimulus (●) attained a VS that was larger than the VSs obtained for other frequencies and was significantly larger than zero (Rayleigh test; P < 0.01), indicating significant phase locking to the stimulus. However, the ability of the neural signal to phase lock to the visual stimulus diminished at repetition rates of 3.5 and 4.25 Hz, a result that was common in the recorded population of neurons. Of 95 sites in which a significant response modulation was induced by low-frequency stimuli (Rayleigh test; P < 0.01), only in 22 sites did the response modulation maintain significance up to stimulus frequencies of 4.25 Hz.

The results of the VS analysis for the bimodal responses are shown in Fig. 7, middle. Only in one case was the auditory and visual stimuli completely synchronized (bold frame in Fig. 7; $f_{\text{vis}} = f_{\text{aud}} = 1.25$ Hz). In this case, the VS of the bimodal response was 0.58, significantly larger (bootstrap; $P < 0.01$) than the corresponding auditory only vector (VS = 0.21) or visual only vector (VS = 0.34). In the other incongruent cases, the VSs elicited by the frequencies of the visual stimuli (Fig. 7, top, ●). The VSs elicited at the frequency of the auditory stimulus (Fig. 7, middle, ○) were not significantly different from the corresponding VSs calculated from visual unimodal stimuli (Fig. 7, bottom, ○). Thus in this site, phase locking to the temporal modulation of the stimulus was better for temporally congruent visual and auditory stimuli than for any of the modalities alone or for incongruent stimulation.

In the following population analysis (summarized in Table 1), we have found that in a little less than half of the multisensory sites (45%) the VS of the response to the congruent 1.25-Hz bimodal stimulus was significantly enhanced (bootstrap, $P < 0.05$) relative to the maximal unimodal VS. In the remaining recording sites, the VS of the congruent bimodal response was not significantly different (25%) or was significantly smaller (30%) than the most effective unimodal VS. This population tendency to enhance phase locking to congruent bimodal stimuli was evident only in the 1.25-Hz tests. For stimulation rates of 2.75 Hz, bimodal stimuli more commonly induced impairment (54% of the population) in the locking to the stimulus rather than an enhancement (18% of the population).

**FIG. 7.** Results of vector-strength (VS) analysis of the responses from a single site. Each plot designates results from a single stimulus combination. Top: the results obtained from responses to the visual alone stimuli (0.5–4.25 Hz); middle: the result obtained from the responses to the auditory alone stimulus (1.25 Hz); middle: the results from the responses to the bimodal combinations of the visual stimuli with the auditory stimulus. Each plot depicts a family of VSs calculated for cycle periods ranging between 0.5 and 10 Hz. ●, the frequency corresponding with the repetition rate of the visual stimulus; ○, the frequency corresponding with the repetition rate of the auditory stimulus. The bold frame contains the results of the bimodal stimulus in which the visual and auditory frequencies were matched at 1.25 Hz.
The distribution of the VS enhancement indices from the congruent 1.25-Hz tests ($n = 175$), was significantly biased toward positive values (Wilcoxon signed-rank test, $P < 0.01$; Fig. 8A). To compare the VS analysis with the SpC analysis, we also calculated the VS enhancement indices excluding the spikes elicited during the period of the first event in the sequence (excluding the period 0–0.8 s). Without this first event, the SpC analysis did not display a population tendency for bimodal enhancement (Fig. 3B). However, the distribution of the VS enhancement indices, calculated from the same spike trains was significantly biased to positive values (Wilcoxon signed-rank test; $P < 0.01$; Fig. 8B). Thus for sequences of congruent bimodal stimuli at a repetition rate of 1.25 Hz, VS enhancement was more widespread than SpC enhancement. However, for higher repetition rates (2.75 Hz), both types of bimodal enhancement were uncommon (Fig. 3D and Table 1).

The results of the VS analysis obtained from the subpopulation of single units ($n = 34$) are summarized separately in Fig. 8, C and D. Both histograms (with and without the 1st event) were significantly shifted to positive values (Wilcoxon signed-rank test; $P < 0.01$). The median of the distribution of the single-unit results including the first event (Fig. 8C) was significantly larger (Mann-Whitney test at a significance level of 0.05 but not 0.01) than the corresponding multiunit results (A). The median of the single-unit results without the first event (Fig. 8D) was not significantly different from the corresponding multiunit results (B).

Results from the incongruent test in which the visual stimulus appeared at a rate of 2.75 Hz and the auditory stimulus at a rate of 1.25 Hz are shown in Fig. 8E and summarized in Table 1. In most recording sites (65%), the bimodal VS was not significantly different from the corresponding unimodal VS. Repeating the VS analysis in other incongruent combinations did not affect the basic result shown in Fig. 8E, that the distribution of the VS indices was not significantly biased to positive values ( Wilcoxon signed-rank test; $P > 0.05$). Because the aim of the following analysis was to understand the VS enhancement, we did not further pursue analysis of incongruent stimuli.

<table>
<thead>
<tr>
<th>Table 1. Comparison of the bimodal VS with the unimodal VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV-$\max$ (A,V), %</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>1.25 Hz</td>
</tr>
<tr>
<td>2.75 Hz</td>
</tr>
<tr>
<td>2.75 Hz (Vis), 1.25 Hz (Aud)</td>
</tr>
</tbody>
</table>

Vector-strength (VS) values are means ± SD.

FIG. 8. Population analysis of VS. A: distribution of the population VS enhancement indices calculated over the complete stimulus period (0–8 s post stimulus) in the congruent 1.25-Hz tests. ■ sites with significant difference between bimodal VS and the maximal unimodal VS [1-tailed bootstrap test; $P < 0.05$]. B: the VS enhancement indices calculated from the same tests as in A without the period of the 1st event (0.8–8 s poststimulus). C and D: histograms showing the VS enhancement indices of the well-isolated single units. The indices calculated over the complete period are shown in C, and the indices calculated without the 1st event in D. Display is similar to A and B. E: distribution of the population VS enhancement indices calculated over the complete stimulus period (0–8 s poststimulus) in the incongruent tests in which the auditory sequence was 1.25 Hz and the visual sequence was 2.75 Hz. Format as in A. F: a scatterplot showing the VS enhancement indices as a function of the corresponding SpC enhancement indices. Indices were calculated from the congruent 1.25-Hz tests.
ing results were based on responses to congruent bimodal stimuli only.

The scatter plot shown in Fig. 8F compares the VS enhancement indices measured from the 1.25-Hz tests with the corresponding SpC enhancement indices. VS indices were measured from responses throughout the stimulus period (0–8 s), whereas SpC indices were measured from responses to the first event. We have found no correlation between the two values. Also, no correlation was found when using SpC indices calculated over the whole stimulus period (not shown). Thus the improvement in phase locking was not dependent on the enhancement in SpCs.

Temporal pattern of the responses

In some of the cases, the VS was enhanced without a substantial increase in the overall number of spikes. Therefore spike count enhancement by itself cannot account for the VS results. Temporal changes in the response pattern have to be taken into consideration as well. What are these unique temporal changes in the pattern of the bimodal response that give rise to the improved VS? To answer this question, we examined the distribution of the spike phases in the stimulus cycle. The three examples in Fig. 9 demonstrate responses to the congruent bimodal stimulus at 1.25 Hz and to its unimodal components. The first example (Fig. 9A) shows results from a site where both the VS and the SpC were significantly enhanced. The second and third examples (Fig. 9, B and C) show results from sites where only the VS was significantly enhanced. In all three examples the visual stimuli (Fig. 9, left) induced spikes that were most likely to occur around a phase of -90° in the stimulus cycle (corresponding to -200 ms). A nonuniform distribution of phases was also observed in the responses to the auditory stimuli (Fig. 9, 2nd column). The phases of the auditory responses however, peaked at a phase of 20–30°, corresponding to -50 ms. Thus auditory responses appeared earlier compared with visual responses, a result that is consistent with the typical delays of visual and auditory responses reported in the OT of barn owls (Gutfreund et al. 2002). Because of this discrepancy in the arrival times (a discrepancy that is expected in natural situations; see DISCUSSION), a simple summation of this discrepancy in the arrival times (a discrepancy that is enhanced. In all three examples the visual stimuli (Fig. 9, left) induced spikes that were most likely to occur around a phase of -90° in the stimulus cycle (corresponding to -200 ms). A nonuniform distribution of phases was also observed in the responses to the auditory stimuli (Fig. 9, 2nd column). The phases of the auditory responses however, peaked at a phase of 20–30°, corresponding to -50 ms. Thus auditory responses appeared earlier compared with visual responses, a result that is consistent with the typical delays of visual and auditory responses reported in the OT of barn owls (Gutfreund et al. 2002). Because of this discrepancy in the arrival times (a discrepancy that is expected in natural situations; see DISCUSSION), a simple summation of the temporal response patterns would lead to a phase distribution that is wider spread in the stimulus cycle, tending to reduce the VS. Interestingly however, in the shown bimodal responses (Fig. 9, 3rd column), the percentage of spikes (relative to total) appearing in between the average phases of the visual and auditory responses (gray area in Fig. 9, A–C), was larger compared with what is expected from simple summation of the spike trains. For comparison the fourth column in Fig. 9 shows the distribution of the phases in a virtual spike train that is the sum of the auditory and visual spike trains. The enlarged concentration of spikes, which is expected to contribute to an increase in the VS, was achieved by either a boost of spikes in the relevant range (star in Fig. 9A) or a reduction in firing outside this range (triangles in B and C). These changes were significant (bootstrap; n = 1,000; P < 0.01, see METHODS for details on the statistical evaluation).

To examine the population tendency, we collapsed the responses to all the events in the 1.25-Hz sequence into one poststimulus time histogram (PSTH) and averaged all recording sites (n = 175). The average PSTHs for auditory and visual stimuli are shown in Fig. 10, A and B, respectively. The average auditory response ascended abruptly, reached a peak ~40 ms after the onset of the stimulus, and declined shortly afterwards. The visual response, on the other hand, ascended gradually and reached its peak 200 ms after the onset of the event. As a result of these characteristic responses, the visual response, on average, overlapped only with the tail of the auditory response. The average bimodal response (shown in Fig. 10, C and D, green curve) was composed of two peaks: the first corresponding with the early auditory response (Fig. 10C, blue curve) and the second corresponding with the late visual response (D, red curve). However, the peak of the late response appeared ~50 ms earlier than the corresponding peak of the visual response, and the tail of the average bimodal response (400–800 ms) was slightly reduced compared with the tail of the average visual response (a reduction that was significant; t-test, P < 0.01).

Figure 10, E and F, compares the temporal shape of the average bimodal response (green lines) with that of the sum of the auditory and visual average responses (magenta lines). The curves were smoothed (5 points moving average), the firing rate at time 0 (baseline) was reduced, and each curve was normalized to the peak. This process was performed separately for the first response (Fig. 10E) and for the subsequent responses (F). In both cases, the average bimodal response was significantly narrower than the sum (bootstrap; n = 1,000; P < 0.01; see METHODS for details on the statistical method), and it peaked at a time not predicted by the summation curve, in between the two unimodal peaks (bootstrap; n = 1,000; P < 0.05 in Fig. 10E and P < 0.01 in F). Thus the observations described in the single examples in Fig. 9 are evident in the population average as well.

The normalized curves of the first event were not qualitatively different from the curves obtained without the first event (compare green curve in Fig. 10, F with E) Neither time to peak nor response duration were significantly different (bootstrap; n = 1,000; P > 0.05), indicating that the temporal features of the bimodal response, unlike the amplitude (see Fig. 10F, inset), were not strongly dependent on the first event. To explore the difference between the 1.25- and the 2.75-Hz tests (Table 1), we averaged the responses of all sites to the single event in the 2.75-Hz sequences and compared with the average responses to the 1.25-Hz sequences (Fig. 11, A–C). The most apparent difference was that the average firing rates were smaller in the responses to the 2.75-Hz auditory, visual, and bimodal stimuli (red curves in Fig. 11, A–C). To graphically compare the temporal patterns of the responses, the 2.75-Hz average responses were normalized to the peak of the corresponding 1.25-Hz responses (superimposed black curves in Fig. 11, A–C). The effect of increasing the repetition rate of the stimulus to 2.75 Hz was a reduction in the response gain of the visual and auditory responses, the temporal patterns were almost identical (Fig. 11, A and B). The effect on the average bimodal response was not only a gain change. The bimodal response to 2.75-Hz stimulus showed the familiar pattern of two peaks (Fig. 11C). However, the second peak appeared at an even earlier time compared with the peak of the average 1.25-Hz bimodal response (bootstrap; n = 1,000; P < 0.05).

What can explain the observation that VS enhancement was common in the 1.25-Hz stimulation rate but not in the 2.75-Hz stimulation rate (Table 1)? One possibility is that the differ-
ences between the responses in the two repetition rates (Fig. 11C) account for this observation. The other possibility is that it is a result of using a shorter cycle period to calculate VSs in the 2.75-Hz tests. To explore this question, we generated 500 triplets of simulated visual, auditory and bimodal spike trains (see METHODS). A VS enhancement index was then calculated for each triplet. Responses were simulated twice, once using the average responses to the 1.25-Hz stimuli (blue curves in Fig. 11, A–C) to estimate the time varying firing rates and once using the average responses to the 2.75-Hz stimuli (red curves in Fig. 11, A–C). The resulting distributions replicated the tendency observed in the data (Table 1). The 1.25-Hz simulations resulted with most bimodal VSs being larger from the unimodal VSs (positive indices), whereas the 2.75-Hz simulation demonstrated the opposite, more negative indices (Fig. 11D). We then calculated the VSs of the 1.25-Hz simulated responses, however, this time for the shorter cycle period of 363 ms (trimming the remaining period from each spike train). This manipulation resulted with a shift of the 1.25-Hz distribution to the negative side, resembling the results obtained for 2.75-Hz responses (Fig. 11E). Similarly, adding a tail of spontaneous activity to the 2.75-Hz simulated responses and calculating the VSs for the longer period of 800 ms resulted with a shift of the 2.75-Hz distribution to the positive side (Fig. 9). Rose diagrams showing the distribution of spike phases (number of spikes found at each phase) in the stimulus cycle (1.25 Hz in all shown cases). The distribution of the phases in the response to the visual stimulus is shown in the 1st column, to the congruent auditory stimulus in the 2nd column, and to the bimodal combination of the 2 stimuli in the 3rd column. The 4th column shows the sum of the auditory and visual distributions minus the level of spontaneous (prestimulus) activity. A: an example from a recording site in which both the VS and the SpC were enhanced. B: an example from a recording site in which only the VS was enhanced. C: an example from a recording site in which only the VS was enhanced. Arrow, the average phase of the bimodal response; black line, the average phase of the visual response; gray line, the average phase of the auditory response. The gray area in the 3rd and 4th columns highlights the spikes that fell in between the average visual phase and the average auditory phase (The percentage of the marked area relative to the total area is written under each diagram). In the 3rd column, the star points to a period in the stimulus cycle characterized by significant enhancement (bootstrap, P < 0.01) in the number of spikes over the sum. The triangles designate areas in the responses where the responses were significantly reduced (bootstrap, P < 0.01) from the sum.
In this case, VS enhancement indices were mostly positive no matter what the stimulus rate was. Thus the negative VS indices commonly observed at 2.75-Hz stimuli were most likely a result of the shorter stimulus cycle relative to the response duration, not reflecting the observed differences between the responses in the two repetition rates. However, differences still remained between the distributions, particularly in the results calculated for a cycle period of 800 ms (Fig. 11F). Although both positive, the median of the 1.25-Hz simulations was significantly larger compared with the median of the 2.75-Hz simulations with the added tail (Mann-Whitney test; \( P < 0.01 \), Fig. 11F). The difference reflects in part the observation shown in Fig. 10D that the average spike rate in the tail of the bimodal 1.25-Hz responses was reduced compared with the visual only responses. This reduction enhanced simulated VS indices so that the distribution was further shifted to positive values. Indeed when simulating with the tail of the 1.25-Hz average responses switched to match the constant rate tail used in the 2.75 Hz, the simulations resulted with a shift of the distribution toward zero (distribution not shown) to better resemble the 2.75-Hz results calculated over a period of 800 ms.

In an additional analysis, we compared the bimodal VS with the VS of the sum of the visual and auditory spike trains (see METHODS). This analysis revealed that in 75% of the congruent 1.25-Hz tests and in 58% of the congruent 2.75-Hz tests the bimodal VS was significantly enhanced (bootstrap; \( P < 0.05 \)) above the VS of the sum (Table 1). Thus bimodal stimuli commonly induced responses that were better phase locked to the stimulus compared with what is expected based on a simple summation of the unimodal responses. This was found for both repetition rates, suggesting that a nonlinear component in the integration process acts to enhance phase locking in both high and low repetition rates. However, in high repetition rates, this effect is masked by the larger reduction in synchrony due to the smaller stimulus cycle (Fig. 11F).

**DISCUSSION**

**Bimodal enhancement**

Our work is the first to characterize multisensory responses in the optic tectum of barn owls. Previous studies in the SC of cats and monkeys established that multisensory neurons tend to respond better to combinations of congruent visual and auditory stimuli (Stein 1998; Stein and Meredith 1993; Wallace et al. 1996). This enhancement was most dramatic when the unimodal stimuli were weak and therefore least effective in driving the response, a principle known as “inverse effectiveness” (Meredith and Stein 1986; Perrault et al. 2005; Stanford et al. 2005). It was argued that this principle is a consequence of statistically optimal integration of visual and auditory inputs (Anastasio et al. 2000). We found that, similar to results in
monkeys and cats, multisensory neurons in the optic tectum of barn owls tend to respond with enhanced firing rates to bimodal stimuli more than responses to the unimodal components of the stimulus. The response of most neurons in our study to bimodal stimuli did not exceed the sum of the unimodal responses (was sub-additive). It should be noted, however, that we did not explore parameters to find the stimulus conditions that give rise to maximal enhancement (Stanford et al. 2005). 1) The visual and auditory stimuli were presented at a fixed range above the unit’s response threshold. 2) We elicited the auditory and visual stimuli simultaneously, a situation argued to be suboptimal for multisensory enhancement (see DISCUSSION on temporal integration in the following text). 3) The visual stimuli were presented from the spatial positions that induced best responses. The auditory stimuli on the other hand were presented in binaural space (ITD and ILD) positions that induced best responses. Because of the strong correlation between binaural space and real acoustic space in barn owls, particularly in the frontal range of space used here (Olsen et al. 1989; Poganiatz and Wagner 2001; Poganiatz et al. 2001) and because of the precise spatial agreement between visual and auditory receptive fields in the OT (Brainard and Knudsen 1993; Knudsen 1982), this procedure ensured that the stimuli were virtually congruent in space. However, it is still possible that aligned free-field auditory stimuli or fully cued virtual space stimuli can give rise to stronger bimodal enhancement compared with the binaural stimuli used here.

Results from studies in cortical multisensory areas revealed multisensory enhancement in some cases and multisensory suppression in others (Ghazanfar et al. 2005; Sugihara et al. 2006). In our data, significant multisensory suppression was observed in some recording sites (Fig. 3, A, C, and E), but these were a minority. Enhancement was predominant in the population. The existence of some multisensory suppression may be accounted for by the fact, mentioned in the preceding text, that we did not use stimuli considered optimal for multisensory enhancement (Meredith et al. 1987; Stanford et al. 2005; Wallace et al. 1996). Nevertheless, we cannot rule out the possibility that in a subpopulation of neurons, suppression of responses to bimodal stimuli is prevalent and that such type of integration may play an important role in visual-auditory processing in the OT of the barn owl.

The difference between this study and most previous studies is that here we explored bimodal interactions in the context of the position of the stimulus in a sequence. Interestingly, bimodal enhancement of SpC was most apparent in the first stimulus of the sequence but less in subsequent stimuli (Figs. 2 and 3). This reduction in bimodal SpC-enhancement was paralleled by a reduction of responses to both uni- and bimodal stimuli. The first stimulus in a sequence was usually more effective in driving the cell than the stimuli that followed due
to response adaptation (previously reported in tectal neurons) (Gutfreund and Knudsen 2006; Reches and Gutfreund 2008). This result seemingly contradicts the inverse effectiveness principle; stimuli that were less effective in driving responses induced less enhancement. However, there is a conceptual difference between a stimulus that is not effective because of adaptation and a stimulus that is not effective because of its low signal-to-noise value: stimuli that are part of a monotonic ongoing sequence provide no new information about the world. Optimal integration principles predict that unimodal stimuli that provide unreliable information will give rise to more bimodal enhancement than stimuli that provide reliable information (Anastasio et al. 2000). Our results therefore may agree with principles of optimal integration and, moreover, suggest a unique role for unexpected stimuli in the integration of visual and auditory inputs.

Tectal neurons are sensitive to novel visual and auditory features (Reches and Gutfreund 2008) consistent with the hypothesis that the OT (SC) is involved in selecting salient targets (Horwitz and Newsome 1999; Maczko et al. 2006). We speculate that visual and auditory information is combined by tectal neurons to enhance the task of detecting salient events. This idea is supported by the stronger bimodal enhancements we have observed in the first “less expected” stimuli, stimuli that are considered more salient (Tiitinen et al. 1994).

Temporal coding

In addition to detecting salient events, it is also important to obtain information about the temporal pattern of the ongoing stimulus. The ability of spike trains to follow the temporal modulation of the stimulus, known as phase locking, allows neurons to relay important information about the stimulus that can be used for cross- and intermodality bindings (Cariani 2004). It was shown that neurons in the midbrain of barn owls are capable of phase locking to the amplitude modulations of auditory stimuli up to rates of several hundred hertz (Keller and Takahashi 2000). On the other hand, visual stimuli tend to be less effective in driving neurons repetitively in this work (Fig. 7) and in other species (Foster et al. 1985). We show that tectal neurons tend to improve phase locking to congruent bimodal stimuli and that this enhancement is independent of the spike count enhancement; in some cases, phase locking improved while the overall number of spikes did not differ from the unimodal response. Moreover, VS enhancement at the population level was apparent during subsequent periods in the sequence (Figs. 8, B and D), periods during which counting the overall number of spikes was not sufficient to detect bimodal interactions (Figs. 3B and 5B). But is this enhanced phase locking to bimodal stimuli utilized by the brain to improve the performance of the animal? The bimodal enhancement was common in low stimulation rates. Amplitude modulations at 2.75 Hz resulted with an opposite tendency, to decrease phase locking to bimodal stimuli (Table 1). Thus if the enhanced phase locking is used to relay more information in bimodal scenes, this additional information is available only at low stimulus frequencies.

Temporal integration

The most apparent characteristic of bimodal environments is the temporal synchronicity of visual and auditory events arising from the same objects. Several lines of research provide physiological (King et al. 1988; Knudsen and Brainard 1991) and psychophysical (Bushara et al. 2001; Sekuler et al. 1997; Slutzky and Recanzone 2001) evidence that temporal cues are of major importance for perceptual binding of visual and auditory scenes. Relying on temporal cues, however, presents an interesting problem: it is known that the visual and auditory systems encode information in different time scales. Auditory signals are generally processed faster and more precisely in time compared with visual signals (Joris et al. 2004). This difference, which stems from differences in the sensory transmission and in the peripheral sensory systems, is apparent in the response properties of tectal neurons of the barn owl (Gutfreund et al. 2002) (Fig. 10, A and B). As a result, a coherent (natural) bimodal scene gives rise to signals that are not coherent in the brain. Moreover, because of the finite velocity of sound compared with light, the temporal relations of visual and auditory signals depend on the distance of the bimodal object from the animal (Alais and Carlile 2005; Spence and Squire 2003). Single-cell study by Meredith et al. (1987) reported that maximal bimodal enhancement is achieved when the peaks of the unimodal activities are temporally aligned, a somewhat paradoxical result because in natural situations, the peaks are rarely expected to align in colliculi neurons. In this study, we synchronized the presentation of auditory and visual stimuli (simulating close bimodal objects) eliciting desynchronized tectal activity. This allowed us to observe that the bimodal response was likely to appear at a temporal window in between the auditory and visual responses (Figs. 9 and 10).

Mechanisms of integration

Multiplication (or coincidence-detection) of the visual and auditory inputs is one way to achieve an enhanced response in the area of overlap. If the activities of the inputs are separated in time and only partly overlapping, as is the case in our study, a multiplication-like interaction can reduce the overall number of spikes while maintaining or improving the phase locking to the stimulus. Our results therefore are in line with the notion that a multiplicative component contributes to the integration process of visual and auditory inputs (Meredith and Stein 1986).

The mechanisms that give rise to such multiplication-like behavior of the output are unknown. It was proposed that N-methyl-d-aspartate receptors presented in multisensory neurons in cats provide a molecular mechanism for multiplication (Binns 1999; Binns and Salt 1996). But, Patton and Anastasio (2003) argued that the experimental results of Binns and Salt (1996) as well as previous results by Meredith and Stein (1986) can be reproduced in a model using simple neural mechanisms such as synaptic summation and threshold, not requiring a specialized synaptic multiplication mechanism. A simple threshold mechanism may explain the results of the present study as well: in the bimodal stimuli, the late visual inputs arrive on top of a raised membrane potential and are therefore likely to reach threshold earlier.

One observation in our results, however, is difficult to explain by a simple threshold mechanism: the late reduction of firing rates in responses to bimodal stimuli observed in some of the recordings (Figs. 9C and Fig. 10D). This observation may
point to the involvement of inhibitory processes in the integration of visual and auditory information.

An alternative hypothesis is that the nonlinearity reported in this study arises from top-down connections that have been shown to dramatically shape multisensory integration in the SC of cats (Stein et al. 2002). Analogous top-down connections have been shown to modulate visual and auditory response properties of tectal neurons in the barn owl (Winkowski and Knudsen 2006, 2007). The influence of such connections on the integration of visual-auditory inputs in the barn owl is currently unknown.

Summary

This study demonstrates that multisensory neurons in the OT of the barn owl tend to respond to a bimodal stimulus stronger than to the unimodal stimuli (multisensory enhancement). However, the results cast two novel constraints on this enhancement: 1) the enhancement was context dependent, occurring primarily in the first stimulus of a sequence and to a much lesser extent in repeated stimuli. 2) The response enhancement was more likely to occur in a period in between the times of maximal auditory and visual responses. Moreover, in some cases, bimodal suppression was observed outside this period. These characteristics allowed better phase locking to repeated bimodal stimuli (at low repetition rates) even without an increase in the overall number of spikes. The findings are consistent with the idea that visual and auditory information is integrated to allow a neuron to better detect an unexpected event as well as to better convey information about the temporal modulation of the stimulus.

Acknowledgments

We thank Prof. Eli Nelken for careful reading of manuscript and F. Milmann for technical support.

Grants

This work was supported by a Focal Initiatives in Research in Science and Technology (FIRST) grant from the Israel Science Foundation and by a Joel Elkes grant from the National Psychobiology Institute in Israel (founded by the E. Smith family).

References


Knudsen EI, Blasdel GG, Konishi M. Sound localization by the barn owl (Tyto alba) measured with the search coil technique. J Comp Physiol A 133: 1–11, 1979.


