Combined Auditory and Visual Stimuli Facilitate Head Saccades in the Barn Owl

(Tyto alba)

Elizabeth A. Whitchurch* & Terry T. Takahashi

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

Running head: Audiovisual integration in the barn owl

*Corresponding author

Elizabeth A. Whitchurch

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

lwhitchu@uoneuro.uoregon.edu

541-346-4544 (phone); 541-346-4548 (fax)
Abstract:

The barn owl naturally responds to an auditory or visual stimulus in its environment with a quick head turn toward the source. We measured these head saccades evoked by auditory, visual, and simultaneous, co-localized audiovisual stimuli to quantify multisensory interactions in the barn owl. Stimulus levels ranged from near to well above saccadic threshold. In accordance with previous human psychophysical findings, the owl’s saccade reaction times (SRTs) and errors to unisensory stimuli were inversely related to stimulus strength. Auditory saccades characteristically had shorter reaction times, but were less accurate than visual saccades. Audiovisual trials, over a large range of tested stimulus combinations, had auditory-like SRTs and visual-like errors, suggesting that barn owls are able to use both auditory and visual cues to produce saccades with the shortest possible SRT and greatest accuracy. These results support a model of sensory integration in which the faster modality initiates the saccade and the slower modality remains available to refine saccade trajectory.
Introduction:

Objects in nature are often both visible and audible, and studies have shown that adding an auditory component to a visual stimulus decreases reaction time (Colonius and Diederich 2004; Corneil and Munoz 1996; Diederich et al. 2003; Engelken and Stevens 1989; Harrington and Peck 1998; Hughes et al. 1994; Miller 1982; Raab 1962; Todd 1912; Welch and Warren 1986) and improves detection and orientation (Corneil et al. 2002; Jiang et al. 2002; Stein et al. 1988; Stein et al. 1989; Wilkinson et al. 1996). How are audition and vision, two peripherally independent senses, integrated to enhance behavior?

A certain amount of improved performance in multisensory tasks may be due to simple probability. In the same way that flipping two coins instead of one increases the chances of obtaining at least one “head”, the probability of either hearing or seeing a weak multisensory stimulus is greater than the probability afforded by either modality alone. Raab (1962) extended this idea of statistical facilitation to reaction times. According to his “race model”, the two sensory signals race along separate pathways toward a common response generation site, and the winning modality triggers the response. If stimulus conditions are such that the two modalities have an equal chance of winning the race, the race model predicts that the multisensory reaction time will be, on average, 0.6 standard deviations earlier than the mean unisensory reaction time (Raab 1962). Multisensory reaction times that are quicker than predicted are said to positively violate the race model (Hughes et al. 1994; Miller 1982). Positive violations have been taken as evidence that instead of remaining separate, the two modalities “converge”, giving rise to reaction
times that are quicker than the race model allows (Arndt and Colonius 2003; Harrington and Peck 1998; Hughes et al. 1998; Hughes et al. 1994; Nozawa et al. 1994). The convergence model is thus distinguished from the race model by its assumption that the two modalities converge to either form a new internal sensory representation, enhance the efficiency of the ensuing motor response, or both.

The study of audiovisual responses in mammalian systems has elucidated three general principles of multisensory integration: for review, (Calvert et al. 2004; Spence and Driver 2004; Stein and Meredith 1993). First, when the auditory and visual components of an audiovisual stimulus are aligned in space, the probability of correctly localizing the target is increased relative to trials in which they are misaligned (Jiang et al. 2002; Stein et al. 1988; Stein et al. 1989). Conversely, spatial misalignment lengthens saccadic reaction times beyond what would be seen with one stimulus alone (Corneil and Munoz 1996; McSorley et al. 2005). Second, auditory and visual information must be aligned in time. If the timing of the auditory and visual stimuli compensates for the difference in processing times between the two modalities, facilitated behavioral (Corneil et al. 2002) or neural responses (Meredith et al. 1987) are more likely. Third, multisensory facilitation is more likely to occur when the unisensory components are presented at low amplitude or signal-to-noise ratio. This has been termed “inverse effectiveness” (Stein and Meredith 1993). Most recently, Stanford et al. (2005) showed that neural responses to audiovisual stimuli were most likely to be super-additive (i.e. facilitated beyond the predicted linear sum of the unisensory components) when near-threshold unisensory stimuli were presented. Similarly, feline detection/orientation behavior in response to
audiovisual stimuli was most enhanced when unisensory stimuli of low efficacies (dim and quiet) were combined (Stein et al. 1989). Moreover, the saccades of human subjects to auditory, visual, and audiovisual stimuli in the presence of background auditory and visual noise were initiated as rapidly with SRTs typical of unisensory auditory trials and high accuracy typical of unisensory visual trials (Corneil et al. 2002). To our knowledge, this tendency to observe the “best” features of acoustically- and visually- guided saccades had not been previously noted in studies using higher signal-to-noise ratios.

The present study takes an ethological approach to multisensory integration, using the barn owl (Tyto alba) as a model. When the owl hears or sees an object of interest, it naturally turns its head to face the object. This stereotyped saccadic behavior, like primate eye movements, can be measured with precision (Knudsen et al. 1979; Knudsen and Konishi 1979). Moreover, the barn owl, renown for its ability to localize sound, has a precise spatiotopic map of auditory space in the external nucleus of its inferior colliculus (ICx) that projects topographically to the retinotopic map in its optic tectum generating a physiologically accessible multisensory map of auditory and visual space (Knudsen 1982; Knudsen and Konishi 1978). The barn owl thus provides us with a model system in which to explore the integration of information from two modalities in the context of its natural behavior.

Here, we measured saccadic reaction time (SRT) and saccadic error in auditory, visual, and simultaneous, spatially aligned audiovisual conditions. The targets spanned a wide range of stimulus locations and strengths, including those unisensory stimuli that were
significantly less effective in eliciting fast, accurate saccades. For most audiovisual combinations, SRTs were no earlier than the earlier of the two modalities, typically, audition, and were no more accurate than the more accurate of the two, typically, vision. However, saccades to audiovisual targets had the short latencies typical of saccades to auditory targets and the accuracy typical of those to visual targets, thus revealing the best features of both modalities, in agreement with a previous human audiovisual study (Corneil et al. 2002). These results are not consistent with the race model which predicts that the winning modality would determine both the SRT and error. Instead, our results support an “updating” hypothesis of sensory convergence in which the first modality to reach threshold triggers the saccade, while information from the second modality is available to update and refine the target position (Corneil et al. 2002; Van Opstal and Munoz 2004).

**Methods:**

All experiments were carried out under protocols approved by the University of Oregon Institutional Animal Care and Use Committee and in compliance with *The Guide for the Care and Use of Laboratory Animals*, (Institute of Laboratory Animal Resources (U.S.) and NetLibrary Inc. 1997).

**Apparatus**

All experiments were conducted in a double-wall, sound-isolating anechoic chamber (Industrial Acoustics Company, IAC), the properties of which have been described previously (Spitzer et al. 2003). The ambient noise inside the chamber was below 18 dB
SPL_A, between 2 and 10 kHz. Head-aim was recorded and sampled at 468 Hz with a custom-built magnetic search coil system (Remmel Labs, Ashland, MA). The head coil was mounted on a post that was cemented to the skull prior to training. The head coil system was calibrated over a ± 50° range in azimuth and elevation at the beginning of each session. This compensated for any day-to-day changes in the imposed magnetic field and enabled the system to remain accurate within 1° to 2°.

**Acoustical stimuli**

One of thirty pre-synthesized, 100 msec auditory stimuli was randomly presented in each auditory or audiovisual trial. Stimuli were broad band noises, flat between 2 and 12 kHz and differed only in fine structure. The noises were digitally synthesized, ramped on and off by a 5 msec cosine envelope, and sampled at 30 kHz (Tucker-Davis Technologies Power DAC PD1). The analog signal was fed to a programmable attenuator (PA4, Tucker Davis Technologies), amplified (HB6, Tucker-Davis Technologies), and presented through one of seven lightweight speakers (Peerless, 5.08 cm cone tweeters). The speakers’ frequency response characteristics were flat (within 3 dB) between 2 and 12 kHz. Auditory levels ranged from -12 dB to 6 dB, in increments of 3 dB on an A-weighted SPL scale. The lower levels were extrapolated from attenuation/SPL relationships recorded at detectable levels with a ½ inch microphone (Brüel & Kjaer model 1760) and sound-level meter (Brüel & Kjaer model 2235).

**Visual stimuli**
The visual stimulus was a single, stationary 100 msec flash from a red (635 nm) light emitting diode (LED), presented in the otherwise completely darkened, light-sealed chamber. Luminous intensities were manipulated by ranging the applied voltage. Six intensities were used (5.2 cd, 0.052 cd, 0.0053 cd, 0.0013 cd, 0.00077 cd, and 0.00031 cd). Each LED was suspended directly in front of one of the seven speakers.

**Behavioral Paradigm**

Two adult barn owls (*Tyto alba*) were hand-reared and trained to make head saccades to auditory, visual, and simultaneously presented audiovisual stimuli. Because the owl’s eyes are relatively immobile (du Lac and Knudsen 1990), gaze was estimated by head-aim. Stimulus synthesis, data acquisition, and data analysis were all carried out with Matlab 6.5 (The Mathworks). Animals were given food rewards and maintained at 85% free-feeding body weight.

Trials were conducted while the bird was tethered to a perch located near the center of an isolated, completely darkened room. Food rewards (mouse bits) were presented from a remotely controlled feeder located in front of the bird, at the level of the perch. Each session, seven targets were pseudo-randomly positioned on either a “T” shaped or “inverted-V” shaped frame, separated by at least 7° (Figure 1A, black and grey orientations, respectively). The frame was alternated between the “T” and “inverted-V” orientations on a session by session basis. Target locations ranged from -30° to 30° in azimuth and 0° to -30° in elevation, with 0° azimuth and 0° elevation being at the intersection of the owl’s midsagittal plane and eye-level.
Birds were trained to localize both auditory and visual stimuli over the course of approximately 4 months. During initial training, only clearly audible and visible stimuli were used (30 dB, 5.2 cd), and the birds were rewarded for localization attempts directed toward the quadrant of space containing the target. As performance gradually improved, the reward criteria became more stringent. Eventually, the bird localized clearly audible and visible stimuli within 5° on more than 90% of the trials. Weaker stimuli (lower SPL or luminous intensity) were then introduced. The training period ended when the accuracy and precision of the birds remained stable for each given modality and stimulus-strength. Data from the training period are not reported here.

The 11 months following training yielded 97 and 94 test sessions for birds ‘N’ and ‘J’ respectively. There were 40 possible rewards within a given session. The session ended when all rewards were dispensed. Sometimes the bird’s performance would warrant a reward on every trial. In these sessions, there were only 40 trials. Other times, the bird’s behavior fell short of the reward criteria and would not be rewarded for up to 15 trials. Such sessions were ended at trial number 65. Analysis included both rewarded and unrewarded trials. In general, rewards were given for head turns terminating within 5° of the target. Trials with very low level stimuli, however, were rewarded when localization attempts were directed in the quadrant of the target, so as to maintain motivation. Unlike earlier head-turn experiments in owls (Knudsen et al. 1993; Poganiatz et al. 2001), we did not require the initial head-aim to be within some radius of a central starting location. Instead, a trial was manually triggered when the bird’s spontaneous head turns ceased and
its head came to rest, typically within 40° of being aligned with its body, as seen with an infrared camera. Post-hoc analysis discarded any trials during which the head was moving faster than 10°/sec prior to stimulus onset, thus avoiding saccades with head motion immediately prior to and during stimulus presentation. The angular distance to the target is expressed relative to the initial head-aim and is referred to as target eccentricity.

Head-aim was recorded for 3 seconds, a time-span comprising a 200 msec pre-stimulus window, the 100 msec stimulus, and a 2,700 msec post-stimulus window. The inter-stimulus interval varied from 15 to 90 sec. A total of 9,872 trials were recorded, 2,779 of which were excluded from analysis, as described below.

Of excluded trials, 66% were those in which the head moved more than 2° in azimuth or elevation in the 200 msec pre-stimulus window, or before 40 msec post-stimulus onset. Head turns initiated before 40 msec post-stimulus onset were almost never directed towards the target and thus not stimulus driven. Trials were also excluded if, at any point during the 3 second recording, the head was outside the calibrated region of our coil system (100° x 100° perimeter, Figure 1A). Seventeen percent of excluded trials were outside of the calibrated region. The remaining exclusions were those trials in which the targets were located above the initial head-aim and/or either too close or too far away; In training and preliminary testing, both birds showed a bias against responding to stimuli closer than 12° and farther than 55° away, as well as those targets presented in the upper visual hemi-field.
The 7,093 accepted trials were categorized as responses or non-responses. Non-responses were relevant in measuring the probability of response, especially in the trials with very low stimulus intensities. Each trial contained an auditory, visual, or audiovisual stimulus, and therefore any head turn with a SRT falling between 40 and 600 msec of stimulus onset was scored as a response. More than 99% of scored responses were directed toward the quadrant of space containing the target. Those trials in which the SRT was greater than 600 msec were instances in which the birds typically did not make any head turn at all and were thus scored as non-responses.

We tested 7 sound pressure levels (SPLs - henceforth also referred to as levels for simplicity) and 6 luminous intensities, resulting in 13 unisensory conditions and 42 audiovisual combinations. For each bird, 70 – 120 trials were recorded for each unisensory condition and 30 – 60 trials for each audiovisual condition. In each session, trials varied randomly in stimulus-strength, location, and modality (auditory, visual, and audiovisual). Each session had a random proportion of unisensory and multisensory trials; however the average session consisted of 30% unisensory and 70% multisensory trials. The luminances, SPLs, and combinations of luminance and SPL also pseudorandomly varied on a session by session basis, and each session included very strong as well as very weak stimuli.

Data analyses

Saccade Reaction Time (SRT)
Figure 1B shows 20 randomly selected, visually evoked, saccades from both *birds ‘N’* and ‘J’. Instantaneous speed from one 40° saccade (Figure 1C) was computed by dividing the distance between consecutively sampled locations by the time between sampling points. SRT was defined relative to stimulus onset as the first point in time at which the instantaneous head speed continuously exceeded two standard deviations of the average speed measured in the 200 msec pre-stimulus window (Figure 1C, open circle).

Preliminary analysis showed a negative correlation between target eccentricity and SRT for auditory trials in *bird ‘N’* (\(r (566) = -0.1228, p < 0.01\)) and *bird ‘J’* (\(r (550) = -0.3126, p < 0.0001\)). However, correlations between SRT and target eccentricity were not significant in visual trials for *bird ‘N’* (\(r (579) = 0.005, p = 0.9135\)) and *bird ‘J’* (\(r (604) = -0.0627, p = 0.1608\)). No significant correlation was found in bimodal trials between SRT and target eccentricity for *bird ‘N’* (\(r (1514) = -0.0087, p = 0.73\)), but this correlation was significant in bimodal trials for *bird ‘J’* (\(r (1558) = -0.1465, p < 0.0001\)). Accounting for this correlation in an analysis of covariance, however, did not affect the overall trends reported below. Thus, trials were pooled across target eccentricities in all further SRT analysis. An analysis of variance (ANOVA) was used to test for main effects of stimulus-strength on SRT in unisensory conditions. Mean SRT changed across conditions in a sigmoidal fashion and was fit by the least squares method for display. *A-priori* pair-wise comparisons (auditory vs. audiovisual, visual vs. audiovisual) of mean SRTs were conducted utilizing the Bonferroni correction for non-orthogonal test design (\(\alpha = (0.05/\text{number of comparisons})\)).
Observed audiovisual SRT distributions were compared to theoretical race model

distribution boundaries computed assuming: 1.) The auditory and visual information
remain separate. 2.) Either the auditory OR the visual stream may trigger a response. 3.)
The first signal to arrive at the motor control center will determine the SRT (Raab 1962).

In addition, any particular model from the class of models based on these basic ‘race-
model’ assumptions will also assume a certain level of unisensory dependence, or
independence (Hughes et al. 1994). Regardless of the level of dependence however, the
upper boundary of SRT facilitation is statistically predicted by the *Miller Inequality*,
which sums the unisensory cumulative probabilities at each SRT (Miller boundary – Eqn.
1, Figure 2, dot-dash line) (Hughes et al. 1994; Miller 1982; Townsend and Wenger
2004):

\[
P(T_{\text{min}} \leq t | \text{AV}) \leq P(T_A \leq t) + P(T_V \leq t), \quad \text{Eqn. 1}
\]

where \( t \) is time relative to stimulus onset, and \( T_A \) and \( T_V \) are times associated with
processing auditory and visual streams, respectively, drawn randomly from the
distribution of unisensory trials. \( T_{\text{min}} \) is the smaller or the two values, \( T_A \) and \( T_V \). If a
significant proportion of the observed multisensory SRT distribution falls to the left of
this upper boundary of the race-model predictions, the distribution *positively* violates the
race model; the response time is shorter than statistically expected, were the birds to act
on the signal contributed from the earlier-arriving modality.
The lower boundary of race-model SRT predictions can be found by using the Grice inequality, which takes the maximum unisensory probability at each SRT value (Grice boundary – Eqn. 2, Figure 2, yellow line) (Grice et al. 1984; Hughes et al. 1994; Townsend and Wenger 2004):

\[ P(T_{\min} \leq t | AV) \geq \max [P(T_A \leq t), P(T_V \leq t)]. \] 

Eqn. 2

Audiovisual SRT distributions showing a significant proportion of their distribution to the right of the Grice boundary are considered examples of negative race-model violations, in which the SRTs are longer than predicted by any version of the original race model.

Race model predictions and observed audiovisual SRT distributions were compared using a one-tailed, one-dimensional, two-sample Kolmogorov-Smirnov test (KS-test) at \( \alpha = 0.05 \). Bonferroni corrections were applied for multiple comparisons. Observed distributions were subtracted from theoretically predicted distributions and plotted as difference graphs (insets, Figures 8, 9). Significantly positive differences between the Miller boundary and observed SRT distributions indicate conditions in which the observed audiovisual SRT was shorter than predicted by the race model. Significantly negative differences between the Grice boundary and observed SRT distributions represent conditions in which the observed audiovisual SRT was longer than predicted by the race model.
Response percentage

Response percentage was defined for each condition as the percentage of accepted trials in which there was an obvious response (see SRT criteria above) toward the target within 600 msec of stimulus onset. The vast majority of trials in which the bird did not respond before 600 msec had no obvious saccade throughout the 3 sec recording period. Theoretical audiovisual response percentages were predicted from a probabilistic combination of the observed unisensory response percentages (Eqn. 3):

$$P_{AV} = (P_A + P_V) - (P_A \times P_V),$$  \hspace{1cm} \text{Eqn. 3}

where $P_{AV}$, $P_A$, and $P_V$ are the probabilities for responding to an audiovisual, auditory, and visual stimulus respectively.

Localization error

The end point of a saccade was designated when the head speed dropped below 5°/sec, (Figure 1). Across saccades, this criterion for head speed accurately delineated the end position of the saccade, (e.g. Figure 1C). Errors in azimuth and elevation were measured, respectively, as the horizontal and vertical distance between the saccade endpoint and the target. Total error was computed by vector summation of the horizontal and vertical components.

ANCOVA’s were conducted to assess total errors co-varying with target eccentricity, as a function of stimulus-strength. Slopes and marginal means for the fitted lines are reported
for each condition. The marginal mean error is the mean of the total error, adjusted for
the linear relationship between total error and target eccentricity. The marginal mean
errors for each condition were computed at a value of 28° from the target, the general
midpoint of tested target eccentricities. Marginal means were plotted against stimulus-
strength and fit by least-squares with a sigmoid. Pair-wise comparisons of marginal
means were conducted for unisensory conditions, with Bonferroni corrections at $\alpha = 0.05$.

**Results:**

We analyzed 7,093 head turns toward auditory, visual, and audiovisual stimuli presented
in the lower frontal hemisphere across a wide range of stimulus intensities. In
audiovisual trials, the light and sound had synchronous onsets and equal durations and
were presented from the same location. Data from the two birds, ‘N’ and ‘J’, are
individually reported and plotted, and the trends described herein are not significantly
different between birds, unless otherwise noted. The statistical significance of each
analysis was determined separately for each bird (Figures 1-10), with the exception of the
multivariate analysis illustrated in Figure 11.

*Responses to Unisensory Stimuli*

Auditory and visual unisensory trials were randomly interleaved with multisensory trials
in each daily test session. The measures of response percentage, SRT, and error were
monitored for each head saccade and individually reported below. The effectiveness of
stimuli at evoking a response of the seven auditory levels and six visual intensities ranged
from near to well above threshold. At the highest auditory and visual stimulus levels, both birds responded at least 99.9% of the time. The weakest auditory stimuli were minimally effective, in that both birds responded to the stimulus only 20% of the time. The weakest visual stimulus evoked a response 75% and 35% of the time for birds ‘J’ and ‘N’ respectively. Predictably, response percentage in unisensory trials increased as auditory and visual stimulus-strength increased (Figure 3).

Mean unisensory SRTs decreased toward an asymptote as auditory and visual stimulus-strength increased (Figure 4A, C, respectively; lighter shades = weaker stimuli). At the highest levels tested, SRTs of auditory trials were significantly shorter than those of visual trials (bird ‘N’: auditory: 72 ± 2 msec; visual: 146 ± 2 msec; t-test, p < 0.001). Figures 4B and 4D show the cumulative probability distributions for auditory and visual SRTs for each condition. For both modalities, the variance, which may be visualized by the slopes and ranges of the cumulative distribution functions, increased for weaker stimuli. Response percentages, taken from the cumulative response probability at 600 msec post stimulus onset, also decreased systematically with stimulus-strength.

Figure 5 shows the analysis of the head-turn accuracy to auditory (A-C) and visual (D-F) targets. Figures 5A and 5D demonstrate that saccade error depended on target eccentricity (angular distance to the target, relative to the initial head aim; See Methods) as well as stimulus strength. We therefore report the marginal means (B, E) and the slopes (C, F) of the regressions shown in Figures 5A and 5D. Just as SRT and response percentage improved systematically with stimulus strength, saccade error decreased as
SPL and luminous intensity increased (B, E). The error for visual trials, however, reached an asymptote more quickly than those of auditory trials for the range of magnitudes tested. The degree to which error depended on target eccentricity also diminished as stimulus strength increased (slope; Figure 5C, F).

To better understand the nature of the errors, saccade total error was divided into its azimuthal and elevational components. Figure 6 shows the azimuthal and elevational components of saccade length for strong and weak unisensory stimuli plotted against target eccentricity. Perfect performance in either dimension would thus yield points along the unity slope line (dashed black line). Stronger auditory stimuli, (black dots, see Figure 4 for relative stimulus effectiveness), consistently resulted in accurate saccades (Figure 6A, B). Interestingly, when auditory targets of small eccentricity and high SPL were presented, this bird had a distinct overshoot in elevation (Figure 6B). The regression line for the azimuthal and elevational components of the saccades to the weaker auditory stimuli (grey circles) tended to have a shallower slope compared to the regression line for the louder condition and to the unity line. As stimulus-strength decreased, this tendency to undershoot was more pronounced for elevation than for azimuth (Figure 6A vs. B, but note differences in scales). Furthermore, the scatter about the regression line was larger for elevation than for azimuth, especially for the quieter sound (Figure 6, r² values in insets). Thus accuracy and precision in both azimuth and elevation were better with louder stimuli.
In contrast, for visual intensities that elicited equivalent response percentages, there was no significant difference between head turns to dim and bright stimuli in azimuth or elevation (compare open vs. closed circles in Figure 6C, D; Note: The dim visual condition plotted here was not the dimmest condition tested). There was, however, a slightly greater tendency to undershoot in elevation than in azimuth. These results are also consistent with the data in Figures 5E – F, which shows a significant increase in error for only the dimmest stimulus (at 0.00033 cd, pair-wise Bonferroni corrected, \( p < 0.05 \)). Thus, while visual SRT increases systematically with decreasing stimulus-strength, visual accuracy remains virtually unaffected.

Responses to Audiovisual Stimuli

Below, we compare results from the multisensory trials with those of unisensory trials. Measured audiovisual response percentages, SRTs, and errors are then compared to race model predictions.

When weaker auditory and visual stimulus conditions were combined in audiovisual trials, the probability of observing a head turn often increased as compared to unisensory levels (Figure 3, grey). Given the auditory and visual unisensory response probabilities, predicted audiovisual response percentages were computed assuming the race model (Eqn. 3) and plotted (Figure 3, black). In general, the observed probability of responding to an audiovisual stimulus was indistinguishable from the theoretical probability of responding to the light, the sound, or both. Only one stimulus combination yielded a
response probability that was noticeably higher than predicted out of 42 conditions for 2 birds (Figure 3A, -9 dB, 3.1x10^{-4} cd, bird ‘N’).

Saccade Reaction Times in Audiovisual Trials

Mean audiovisual SRTs typically followed the shorter of the two unisensory SRT means. In Figure 7A – F, the mean SRT (± SEM) is plotted against SPL for auditory (short-dash) and audiovisual trials (solid). The long-dashed horizontal line in each graph indicates the mean visual SRT (± SEM) for the luminance shown. The figure thus represents 42 audiovisual combinations (7 auditory; 6 visual). When the visual stimulus was dim, as in Figure 7A, most auditory SRTs were shorter, and the audiovisual SRTs closely approximated the auditory SRTs. With brighter visual stimuli, the audiovisual SRTs followed the visual SRT values at low SPL combinations and the auditory values at higher SPL combinations (Figure 7D – F). There was one condition for each bird in which the mean audiovisual SRT was significantly shorter than the shortest unisensory SRT (black arrow, Figure 7E). All of these results are qualitatively consistent with the race model. On the other hand, one stimulus combination in each bird (example bird ‘N’ grey arrow, Figure 7F, bird ‘J’ same stimulus combination, data not shown) resulted in a mean audiovisual SRT that was significantly longer than the shortest unisensory SRT. The following distribution analysis addresses whether these differences in mean SRT significantly violate race-model predictions.

Figure 8 shows cumulative probability functions for audiovisual trials at a single low level SPL (-9 dB) and six luminances (A – F) for bird ‘J’. Under these conditions, the
distribution of auditory and visual SRTs often overlapped. For conditions in which the visual stimulus was fairly dim (Figure 8A – D) the effectiveness of either unisensory stimulus alone was less than optimal in evoking a quick, reliable response. Thus, these combinations theoretically provided the best opportunity for observing positive race-model violations. The insets plot the difference between the upper boundary of race model predictions (Miller boundary: see Methods) and observed cumulative probability functions. Upward excursions in the inset graphs indicate positive distribution deviations from the upper boundary of the race model. The statistical significances of the excursions were computed with the KS-test (see Methods). As exemplified in this set of conditions, all observed SRTs were consistent with the predictions of the race model, and no significant positive race violations were found for bird ‘J’. There was only one audiovisual combination in one bird (-9 dB and 5.2 cd, bird ‘N’) in which the SRT was significantly shorter than race-model predictions, and thus qualified as a positive violation. This condition is shown in Figure 9B (Note: the inset difference graph for this figure is computed using the Grice boundary, but ‘#’ marks a significant positive violation, as compared to the Miller boundary).

Negative violations, where the observed SRTs were significantly longer than race-model predictions, were somewhat more frequent and found with brighter (5.2 and 0.052 cd) and louder (-3 through 6 dB) stimulus combinations. The condition marked by the grey arrow in Figure 7F is an example of a negative violation severe enough to significantly affect mean values. There were multiple conditions, however, that showed significant negative violations in the distribution analysis. Figure 9 plots SRT cumulative
probability distributions for a single high-intensity visual stimulus (5.2 cd) and seven SPLs (A – G). As shown in Figures 9D – G, four of the loud/bright combinations resulted in distributions that fell to the right of the Grice boundary (Data shown for bird ‘N’, bird ‘J’ gave similar results.)

Saccade Error in Audiovisual Trials

The race model is based on the competition between two separate streams of sensory information. Does the modality that “wins” the race determine all characteristics of the ensuing saccade, or does the presence of the stream that “lost” have an influence? Specifically, do audiovisual saccades with auditory-like SRTs also have auditory-like errors? In this section, we compare audiovisual errors with unisensory error in the context of race-model predictions. Error and SRT are then considered simultaneously for one exemplary audiovisual combination.

Figure 10 plots marginal mean error against SPL for all 42 stimulus-combinations. The conventions are the same as those of Figure 7. The insets show the slope of the regression between total error and target eccentricity (see Figure 5). The stimulus combinations shown in Figure 10A – C typically yielded audiovisual SRTs similar to auditory SRTs (see Figure 7A – C). Yet, even in these cases where the auditory stream presumably “won” the race to trigger a saccade, the audiovisual errors approached visual values. This would not have been the case were the winning modality to control all aspects of the saccade.
When brighter stimuli (Figure 10E – F) were paired with sounds that were less than -5 dB SPL, visual SRTs were shorter than the auditory SRTs (see Figure 7E – F). For these combinations, the audiovisual errors were, again, closer to the visual errors. In bird ‘N’, whose results are shown, the audiovisual error was slightly but significantly larger than the visual error in 4 of the 42 combinations tested (“#” in Figure 10C, E, F). In bird J, the audiovisual errors were indistinguishable from visual errors in all conditions. As seen with SRT, audiovisual accuracy was never better than that seen in the most accurate unisensory condition (t-test, Bonferroni corrected, p > 0.05, birds ‘N’ and ‘J’).

The results above are inconsistent with the idea that the winning modality in the race model controls all aspects of the saccade. To test this more directly, we compiled a theoretical distribution of audiovisual saccades for each stimulus combination with a Monte Carlo simulation based on the race model. This simulation utilized the basic race-model assumptions: 1) the two modalities raced to the saccade generator, 2) the winner of the race controlled both the SRT and the error of the theoretical audiovisual saccade. For simplicity, we assumed independence of SRTs between the two unisensory streams. For each multisensory combination, one randomly selected auditory saccade SRT was compared to one randomly selected visual saccade SRT. The characteristics of the saccade with the shortest SRT (including error and SRT) were then pooled in a theoretical audiovisual distribution. If the two unisensory SRTs were equal, a modality was randomly selected and the corresponding saccade SRT and error was added to the predicted audiovisual distribution. Both auditory and visual saccades were then replaced in their original unisensory pools and the selection process was repeated 100 times. The
SRT and error of the resulting pool of 100 “race-winning” auditory and visual saccades were analyzed as a theoretical audiovisual distribution.

As expected, the Monte Carlo simulation resulted in audiovisual SRTs and marginal mean errors that reflected the winning modality. For many audiovisual combinations, the errors predicted by the simulation were significantly greater than those observed in the actual audiovisual trials (7/42 bird ‘N’, 9/42 bird ‘J’, t-test, Bonferroni corrected, p < 0.05).

An example of an audiovisual condition in which the observed error was less than that predicted by the Monte Carlo simulation is illustrated in Figure 11 (-3 dB; 5.3e-3 cd, birds ‘N’ and ‘J’ combined). These stimuli were of moderate effectiveness; Unisensory trials with these stimulus-intensities individually evoked a response more than 90% of the time, however both auditory and visual mean SRTs were significantly longer than those seen at the highest intensities (see Figure 4). Figures 11A and B show the distance traveled in azimuth and elevation as a function of target eccentricity in azimuth and elevation, respectively. In azimuth (A), the unisensory and audiovisual trials had similar errors (deviation from the unity line), although the auditory trials had slightly more scatter at the larger target eccentricities than did the visual trials. In elevation (B), the auditory trials had appreciably more scatter than the visual trials, and notably, the errors in the audiovisual trials were more like those in visual trials.
Total audiovisual errors also closely followed the visual error trends (Figure 11C), with the slope and marginal mean error of audiovisual trials being significantly different from auditory but not visual values (pair wise comparison, Bonferroni corrected, \( p < 0.05 \), *birds 'N' and 'J'). However, KS-tests of the SRT cumulative probability distributions, shown in Figure 11D, indicate that the audiovisual SRT was significantly different from the visual distribution, but not the auditory (Bonferroni corrected, \( p < 0.05 \), *birds 'N' and 'J'). Figure 11E is a bivariate summary of these findings, plotting each head turn’s SRT (abscissa) and total error (ordinate). The plot suggests that the audiovisual distribution has a shorter SRT than the visual distribution and a smaller error than the auditory distribution. The bivariate distribution of the audiovisual trials was significantly different from both the auditory and visual distributions (2-dimensional KS-test, \( p < 0.025 \), *birds ‘N’ and ‘J’).

Figure 12 is a summary figure showing mean SRT and error across all unisensory and multisensory conditions for each bird (*N*: panels A-F; *J*: panels G-L). In each panel, the data are normalized to the mean visual values (V-norm, grey circle, dashed lines), at the specified luminous intensity. This normalization was done to simplify comparisons across modalities and stimulus-strengths. The same unisensory auditory data (grey diamonds) are reproduced in every panel for comparison to audiovisual data (black squares), and the SPLs are labeled in panels F and L.

Examination of the unisensory auditory data in Figure 12 shows that as SPL decreased (grey diamonds, left to right), both errors and SRTs increased. When visual stimuli were
added to the sounds at the same SPLs (black squares), however, the errors remained similar to those obtained in visual trials. The vertical disparity between the grey diamonds and corresponding black squares represents the improvement in accuracy afforded by the addition of a visual stimulus. This improvement was more apparent in trials with low-level auditory stimuli than in those with high-level auditory stimuli. For example, in each panel, the greatest vertical separation between black squares and grey diamonds occurs at the right of the plot. This is consistent with the principle of inverse effectiveness: The benefit of adding a second modality (in this case, vision) is most apparent when the auditory component is near threshold.

SRTs also benefited from the inclusion of a second modality in a way that is consistent with the principle of inverse effectiveness. In each panel of figure 12, the horizontal disparity between the black squares and V-norm represents the improvement in SRT afforded by the addition of an auditory stimulus. Across the auditory conditions, this improvement in SRT is more evident in low-intensity visual combinations (A-C, G-I) than high-intensity ones (F, L). For instance, panels B and H represent conditions with lower-intensity stimuli, and each show four audiovisual combinations with SRTs less than half the value of the visual SRT (0.5 on x-axis). Audiovisual combinations incorporating the highest-intensity stimuli, as shown in panels F and L, have SRTs that cluster to the right of 0.5 on the x-axis. Therefore, the shortening of SRT afforded by adding an auditory stimulus to a visual stimulus is most apparent when the visual stimulus is weaker.
Finally, figure 12 demonstrates that when a very strong stimulus is paired with a very weak one, the audiovisual SRTs and errors *both* take on the values of the stronger modality. This can be seen for high luminance, low SPL combinations in panels F and L, by comparing the right-most black squares to their unisensory counterparts. In these cases, the audiovisual responses resembled the visual responses in both SRT and error. Conversely, when the unisensory strengths in the audiovisual combinations were reversed, as in the left-most conditions in panels A and G, audiovisual responses resembled the auditory in both SRT and error. When *both* modalities in an audiovisual combination were strong, as in the left-most squares in panels F and L, the audiovisual error can not be classified as being more like auditory or visual because the owls localized higher SPL sounds with near-visual accuracy. However, the audiovisual SRTs did seem to be more like the shorter auditory SRTs. Therefore, any combination incorporating strong unisensory stimuli produce very little evidence of audiovisual integration.

Across stimulus conditions, the data exemplify the general definition of the inverse effectiveness principle; a best-of-both-worlds audiovisual integration is *most* apparent when stimuli are near behavioral threshold. When such stimuli are used, it becomes clear that the owl’s saccade system uses information from both streams to optimize orienting behavior.

**Discussion:**
Events in nature often have auditory as well as visual components, and given that their signal strengths may vary widely, we examined how a behavioral response might benefit from various combinations of SPLs and luminous intensities. Initial characterization of the barn owl’s saccades to unisensory targets showed that when unisensory signal strength increased, response percentage increased, while SRT and error decreased. Neither the SRT nor the accuracy of an audiovisual saccade was significantly facilitated beyond the level predicted by the race model. However, multisensory saccades had reaction times typical of the auditory saccades while maintaining the high accuracy characteristic of visual saccades. We also found that for stimulus combinations employing bright stimuli, SRTs were longer than the race model’s predictions. These observations argue against the race model as being the sole mechanism of audiovisual integration and suggest a convergence of the auditory and visual modalities within the saccadic sensory/motor pathway.

**Probability of Response:**

Audiovisual response probability was indistinguishable from the predicted probability of responding to vision, audition, or both. To our knowledge, most published psychophysical studies of multimodal integration tend to employ stimuli at amplitudes well above behavioral threshold, thus ensuring detection of nearly all of the stimuli, regardless of modality. By contrast, the behavioral studies of Stein and colleagues did test peri-threshold stimuli, and is therefore relevant. They showed that cats, trained to approach a visual stimulus and press a bar directly beneath the source, were more likely to orient correctly towards the light when a sound accompanied it (Stein et al. 1988;
Stein et al. 1989). The percentage of correct trials was, furthermore, enhanced beyond statistical predictions (Burnett et al. 2004; Stein et al. 1988; Stein et al. 1989; Wilkinson et al. 1996). Although our study included stimuli of comparable magnitudes, it is not directly comparable with these studies of the cat because Stein and colleagues required not only a reaction to the stimuli, but also that the response be to the correct location. By contrast, we considered all saccade-like movements (within criteria; see Methods), regardless of accuracy and precision. The question remains, whether the feline detection of stimuli regardless of correctness would be enhanced to levels beyond that predicted by a reaction to either modality alone or to both.

**Saccade Reaction Time:**

Because saccades to audiovisual targets usually had shorter, auditory-like SRTs, our results are consistent with a number of studies suggesting that a visual SRT will be facilitated by the addition of an auditory component (Arndt and Colonius 2003; Colonius and Arndt 2001; Corneil and Munoz 1996; Corneil et al. 2002; Engelken and Stevens 1989; Frens et al. 1995; Harrington and Peck 1998; Hughes et al. 1998; Hughes et al. 1994; Kirchner and Colonius 2004; Nozawa et al. 1994). Performance meeting this broad definition of enhancement is also predicted by the race model, and our results did not show substantial evidence of SRT facilitation beyond race-model predictions, in contrast to previous results (Arndt and Colonius 2003; Harrington and Peck 1998; Hughes et al. 1998; Hughes et al. 1994; Nozawa et al. 1994).
Instead, the lack of SRT facilitation beyond race model predictions in the owl is consistent with a recent study on multisensory integration in human saccadic behavior (Corneil et al. 2002), wherein SRTs to simultaneous, spatially-aligned audiovisual stimuli were not enhanced beyond race-model predictions. There are three main similarities between the study by Corneil and colleagues and our own that may help explain the lack of SRT facilitation beyond race predictions. First, both studies implemented a “divided attention” task, in which subjects are instructed or trained to localize targets of one modality while ignoring the other. In focused attention tasks, by contrast, subjects are instructed to ignore one modality and must therefore identify the modality of the target before initiating a saccade. Because our divided attention task did not require the modality identification step, unisensory responses and thus predictions made by the race model were quicker, making the observation of positive race violations less likely (Corneil and Munoz 1996). Second, the difficulty of the task was set at a high level in both sets of studies by presenting stimuli with a low signal to noise ratio. Third, for any given trial, the stimulus-strength, location, and modality were randomly selected, and thus unpredictable. Previous studies have linked response facilitation to a priori knowledge of stimulus characteristics (Kirchner and Colonius 2004; Mordkoff and Yantis 1991; Schwarz 1996).

SRTs are also known to be affected by the temporal alignment of the auditory and visual components in a multisensory target (Corneil et al. 2002; Diederich and Colonius 2004). Corneil and colleagues (2002) demonstrated positive violations of the race model when a light led the sound by 100 ms. Because SRTs in unisensory auditory trials of that study
were on average 100 ms earlier than those in visual trials, presenting the light 100 ms earlier than the sound would presumably align the visual and auditory information streams within the saccade-control circuitry. Electrophysiological evidence from the mammalian superior colliculus suggests that this is the condition in which positive violations of the race model are most likely to occur (Meredith et al. 1987). Although the visual and auditory streams were similarly aligned in our study by ranging SPL and luminous intensity (Figures 4, 7), we did not see significant positive race model violations. It is worth noting, however, that aligning the unisensory components in this way during multisensory trials changes not only the processing time, but the strength of the inputs. Future studies in which the temporal alignment of auditory and visual information is achieved by asynchronous presentation of the stimuli, independently of stimulus strength, may yet reveal positive race model violations in the owl’s saccade control system.

Negative Race Model Violations in Saccade Reaction Time:

SRTs that were longer than predicted by the race model were more frequent in our study than in earlier studies of humans. These negative violations were consistently observed when the strongest visual stimulus was utilized in the audiovisual combination (Figure 9). Interestingly, when these bright visual stimuli were presented alone, the SRTs were still longer than those from auditory-alone trials, and therefore, the visual information stream would not have been expected to win a race to saccade initiation in an audiovisual trial. Nevertheless, a significant proportion of SRTs in both birds had vision-like values that skewed the overall multisensory distributions to the right of the Grice boundary. Data
suggesting negative violations have also been reported in focused attention tasks where human subjects are instructed to generate ocular saccades to stimuli of one modality while ignoring another (Hughes et al. 1994). Negative violations can not be explained solely by a race between separate sensory streams (Grice et al. 1984; Hughes et al. 1994).

Negative violations suggest that under certain circumstances, the process of multimodal integration incorporate factors other than timing, such as the strength or precision with which the stimulus is represented within the nervous system. When the two information streams arrive within a time window, the “stronger” or more “focal” representation may control saccade initiation, even though it was not the first to arrive at the saccadic trigger site. It is premature, however, to propose the neural mechanism by which the quality of the neural representation affects SRTs. The role of stimulus salience would be better tested by manipulating stimulus onset asynchrony independently of amplitude.

Negative violations may also be due to a perceived spatial misalignment of the auditory and visual sources. Previous studies have shown that SRTs to targets in the presence of a spatially-misaligned second stimulus are significantly longer than in trials with only one stimulus (Corneil et al. 2002; McSorley et al. 2005; Spitzer and Takahashi 2006). In our study, the LEDs and speakers were co-localized. However, in trials where a quiet sound was presented alone, we saw that the owls’ saccades typically fell short of the target (Figures 5, 6). No such hypometria was observed for visual-alone trials. This difference in saccade metrics in auditory and visual trials may suggest that the neural representation of a quieter sound is shifted toward the midline relative to the visual representation. In
other words, although the LEDs and speakers were aligned in space, the visual and auditory neural images may have been misaligned, effectively forcing the owl into a spatially-misaligned condition, which is known to lengthen SRTs.

The lengthening of SRT due to misalignment may stem from a network of local inhibitory neurons in the superior colliculus (Honda 2005; Munoz and Istvan 1998; Olivier et al. 1999). In general, these studies have shown that the presence of a misaligned distracter decreases buildup of activity in bursting superior colliculus neurons prior to saccade initiation (Olivier et al. 1999), and also lengthens SRTs beyond what is seen for either modality alone (Munoz and Istvan 1998). The negative violations seen here, however, are due to a skew of the distribution toward values more common in visual trials, rather than a shift of the mean toward these longer values. Whether or not lateral inhibition may still play a role in the negative violations observed with our paradigm is empirical, and should be addressed in future electrophysiological studies.

Finally, prior experience may also have played a part in inducing the negative violation. The present study incorporated an accuracy criterion for food rewards. Although owls are renown for their ability to localize sounds, their ability to localize a visual stimulus across intensities is better still, as measured by error. When auditory stimuli were quieter, there was a lesser probability of reward, even with our relaxed criteria in these cases (see Methods). Therefore it is possible that when a visual stimulus was added, the bird weighed the visual input more heavily because it allowed the bird to perform to criterion and receive rewards more consistently. Thus in some or all of the stimulus
combinations, we can not rule out the possibility that the birds may have been
incorporating an additional decision step in the planning the saccade.

*Error:*

Audiovisual error at different stimulus-strengths mirrored the more accurate of the two
unisensory trends and did not improve beyond the level of the most accurate/precise
modality, usually vision (Figure 10). Others have attributed this lack of enhancement to a
ceiling effect, arguing that vision provides such accurate localization that the addition of
auditory information cannot further improve performance (Corneil et al. 2002; Hairston
et al. 2003; Welch and Warren 1986). Our study, however, extended the stimulus range
to include visual trials with intensities so low, that errors were significantly larger than
the asymptotic value (Figure 5), i.e., accuracy was below the hypothetical ceiling, leaving
room for improvement. Adding an auditory component to these dim stimuli however, did
not improve accuracy beyond the level of the better modality (Figure 10A). An absolute
accuracy ceiling, therefore, can not entirely explain the lack of accuracy/precision
enhancement beyond unisensory levels in the owl.

We also compared observed audiovisual errors with the prediction that the winning
modality of the race to saccade initiation determined other features of the ensuing
saccade, i.e. accuracy. Interestingly, when an auditory stimulus evoking
characteristically shorter SRTs was paired with a visual stimulus evoking
characteristically more accurate saccades, the ensuing audiovisual saccade was generally
initiated as early as the auditory saccades and terminated as accurately as the visual. This
observation is inconsistent with the idea that when an object is audible and visible, the owl acts on either the visual or auditory component to generate a saccade, as is the basic assumption of the race model. Our results support Corneil et al. (2002) in suggesting that an audiovisual saccade is often neither purely auditory nor purely visual (whichever won the race) but an optimal combination of both. Furthermore, this effect was seen primarily for stimulus combinations in which neither unisensory component was loud or bright. This is consistent with the inverse effectiveness rule of multisensory integration. Thus, the two modalities do not function independently in the multisensory trial, but rather, are ultimately integrated to enhance the behavioral response.

Neural Mechanisms:

Studies of the mammalian superior colliculus have typically shown that auditory and visual stimuli co-localized at the center of a neuron's spatial receptive field evoke responses that are greater than the response to either modality alone (Frens and Van Opstal 1998; King and Palmer 1985; Meredith and Stein 1986; Populin and Yin 2002; Stanford et al. 2005; Stein and Meredith 1993; Wallace et al. 1998). Although there is some controversy whether the audiovisual response is greater than the sum of the responses to the visual and auditory components (Populin and Yin 2002), this increased neuronal response parallels the improved performance in the detection/orientation studies cited above (Stein et al. 1988; Stein et al. 1989). Stein and colleagues have pointed out that in both neuronal and behavioral responses, multisensory advantages are greatest when the stimuli are weak, i.e., the principle of inverse effectiveness applies. Moreover,
the proper alignment of the stimuli in time and space are required to achieve the improved performance in behavior and the increased response in the neurons.

The recent study of Bell and colleagues (2005), which examined spike timing in addition to spike rate in the superior colliculus of the behaving monkey, sheds light on the improvement of SRTs observed with bimodal stimuli. They showed a significant correlation between the latency of the first spike evoked by weak stimuli and mean SRTs. Importantly, the addition of a second modality decreased both the first-spike latency and SRT. It is difficult, however, to relate these findings to the current study. The paradigm implemented by Bell and colleagues (2005) was a focused attention task in which the visual signal was always the target and the auditory signal was either a distracter or an enhancer, depending on spatial orientation. The auditory stimulus was never presented alone as a target. Therefore, the decrease in SRT could only be compared to the visual-alone trials and not to race model predictions. Likewise, the decrease in first spike latency was again relative to visual-alone trials, even though a number of the cells pooled into the analysis were responsive to the auditory stimulus. Our behavioral results predict that the latency of the first spike evoked by AV stimuli will be similar to those evoked by the auditory component alone, if indeed, first spike latencies of tectal neurons determine SRTs.

The observation that the owl’s saccades have the combination of speed and accuracy characteristic of the auditory and visual modalities respectively suggest an avenue of exploration of neural mechanism in addition to spike rate and latency. This approach
would examine the hypothesis that the first modality triggers the saccade, and the other refines the target location sometime before saccade termination (Corneil et al. 2002; Van Opstal and Munoz 2004). In the initial phase of this “updating” model of sensory convergence, separate auditory and visual sensory responses race to a common saccade generation site, such as the mammalian superior colliculus or avian optic tectum. Then, if the auditory sensory representation brings the saccade generation mechanism to threshold before the visual, the SRT assumes an auditory-like value. Once the saccade is triggered, the target location may be updated and refined by newly arriving visual information. The updating model thus stated, however, does not address negative violations, which may require assumptions about perceived location as well as the incorporation of themes from decision theory, such as prior experience and stimulus salience (Carpenter and Williams 1995).

Human psychophysical experiments have shown that when two targets are presented in sequence, we are able to update the trajectory of a saccade even after the response has begun (Vliegen et al. 2004). Behavioral experiments on the precedence effect demonstrate that such mid-course modification of trajectories is also possible in owls. The precedence effect comprises a cluster of phenomenon related to spatial hearing in echoic environments in which the location of a sound coming directly from the source dominates perception, and an echo, which arrives later, is poorly localized if at all, for review (Litovsky et al. 1999). Owls turn their heads toward the leading source when the delay between the leading and lagging sounds is less than about 10 ms, but begin to localize the lagging sound when the delay is more than 20 ms (Keller and Takahashi
In many of the trials at delays > 20 ms, owls first turned their heads toward one source and then abruptly changed their head trajectory to localize the later source. Evidence for updating was also observed in the course of the experiments conducted in the present study: In those trials where stimuli were inadvertently presented during a “spontaneous” head turn (66% of excluded trials, see Methods), the owls often made appropriate mid-course corrections. The owl is therefore neither blind nor deaf during a saccadic head turn, and the sensory information arriving later can affect the final gaze position.

Neurophysiological studies also suggest that an updating model of sensory convergence is plausible. In single multisensory units in the intermediate and deep layers of the barn owl’s optic tectum, the auditory first-spike latency is approximately 10-20 msec, whereas the first-spike latency to a light flash is approximately 50-80 msec (Knudsen, 1982; unpublished observations). Although the visual and auditory receptive fields generally overlap in space, the visual receptive field is considerably finer than the auditory receptive field (Knudsen, 1982, unpublished observations). The timing and spatial resolution of responses in the optic tectum are therefore consistent with the idea that audition often triggers the saccade while later arriving visual cues refine it. To further study the neuronal correlates of the updating model, it will be necessary to observe the complete sensory receptive field as it evolves over the course of an audiovisual stimulus. This may be possible in the barn owl, wherein computer-synthesized visual stimuli and virtual auditory space techniques now allow the rapid assessment of spatial receptive fields (Keller et al. 1998).
Acknowledgements:

Many thanks to Drs. Michelle Gaston, Kip Keller, Brian Nelson, and Matthew Spitzer for their critical evaluation of this manuscript. We also gratefully acknowledge Dr. Shin Yanagihara for his initial work in training *birds* ‘N’, and ‘J’.

Grants:

This work was supported by grants from the NIGMS (T32-GM07257; EAW), and the NIDCD (DC-03925).
References


Figure 1. Head saccade measurements.
A) Targets were presented in the lower hemi-field on either a frame shaped like a "T" (black) or an inverted "V" (dark grey). The circles represent speakers, each with an LED at its center. The speakers are shown on the T-shaped frame, but were also placed on the inverted-V frame on different days in order to increase the uncertainty about speaker location from session to session. During each session, 7 targets were arranged within 30 deg of the origin (i.e., the apex of the inverted-V frame, or the intersection of the stem and arms of the T frame) and separated by at least 7 deg. All trials accepted for analysis were triggered and terminated within the calibrated region, which subtended 50 deg from the origin (full area of the graph). B) Example saccades, from trials with bright visual stimuli, are shown, (data from bird 'N' and bird 'J'). Saccade eccentricity is plotted relative to initial head position. Light grey lines show distance between saccade end-position and target. C) A typical 40 deg head saccade is plotted in black, (data from bird 'J'). Latency points (circle) and end points (square) were determined by velocity criteria (see Methods).
Figure 2. Cumulative probability distributions predicted by the race model. The dot-dashed line shows the Miller boundary, which is the upper boundary of any race model prediction. When audiovisual SRT distributions are significantly to the left of the Miller boundary, SRTs are generally shorter than predicted by the race-model (dark grey area). The yellow line shows the Grice boundary, which is the lower boundary of the race model. The race model is said to be negatively violated when the observed audiovisual SRT distributions are significantly longer than race model predictions (light grey area). Actual auditory and visual SRT distributions are shown for near-threshold conditions from bird 'J'.
Figure 3. Effects of stimulus strength on response percentage.
Data from the two birds are shown (bird 'N', A&B; bird 'J', C&D). Response probabilities obtained in auditory (white bars) and audiovisual (grey bars) trials are plotted for the SPLs tested at the two weakest luminances tested (cross-hatched bars, 0.00031 cd, A&C; 0.00077 cd, B&D). The black bars represent the predictions of the race model. Observed audiovisual response percentage was either equal to or greater than unisensory values. Observed response percentage generally followed race model predictions for most stimulus pairs.
Figure 4. Effects of stimulus strength on saccade reaction time (SRT).
Mean (± SEM) auditory SRTs from *bird 'N'* are shown (A) to decrease with increasing SPL, (Data shown *bird 'N'*: F(6,560)=84.69, p<0.001; Data not shown *bird 'J'*: F(6,526)=73.49, p<0.001; ). Mean visual SRTs (C) also decreased with increasing intensity, (*bird 'N'*: F(5,466)=188.07, p<0.001; *bird 'J'*: F(5,487) = 21.14, p<0.001; ). Cumulative probability distributions obtained in auditory (B) and visual (D) trials show that distribution variance increased and the probability of response decreased stimulus strength decreased. The response percentage is the cumulative probability at 600 msec after stimulus onset, multiplied by 100.
Figure 5. Effects of target eccentricity and stimulus strength on saccadic error.
Panels A and D show that auditory and visual saccades to more distant targets had greater errors (solid lines: regression fits). An ANCOVA analysis, with initial distance to the target (eccentricity) as a covariate, verified that error depended on stimulus strength for both auditory and visual conditions (Data shown bird 'N' -- Aud: F(6,560)=31.97, p<0.001, Vis: F(5,466)=3.8, p<0.01; Data not shown bird 'J' -- Aud: F(6,526)=22.98, p<0.001, Vis: F(5,487)=7.47, p<0.001). Marginal mean errors were computed for the mean target eccentricity (vertical line at 28°) and plotted in B&E. Auditory marginal error (B) decreased as a function of increasing SPL (±SEM). Only the dimmest visual stimulus produced a significantly different marginal error (E. pair wise comparison, Bonferroni corrected, p<0.05). All but the dimmest visual conditions gave rise to saccades that were more accurate than those of the auditory conditions (t-test of marginal means and slope, Bonferroni corrected, p<0.05). The slopes from the regression fits shown in A&D are plotted in C&F. The degree to which total error depended on target eccentricity decreased as a function of stimulus strength (±SE of slope). For the majority of conditions, visual saccades were less dependent on target eccentricity.
Figure 6. Effects of stimulus strength on azimuthal and elevational components of error. A&B) Saccade length (angular distance) is plotted as a function of initial distance to target (target eccentricity) for relatively loud (6 dB, black solid) and quiet (-6 dB, grey open) auditory trials. Perfectly accurate azimuthal or elevational components of a saccade fell along the unity line (dashed line). Regression lines (solid lines) show that in quieter auditory trials, under-shooting in both dimensions was common. This undershooting was exaggerated in the elevational component and more prevalent for larger target eccentricities. C&D) No significant difference was found between error trends of saccades to bright (5.2 cd, black solid) versus the dimmer (7.7e-4 cd, open grey) visual stimuli. Data from bird 'J'.
Figure 7. Average audiovisual SRTs compared to unisensory values for bird ‘N’. One mean visual (long dash) and all seven mean auditory (short dash) SRTs (± SEM) are shown for each plot as a reference for the audiovisual conditions (solid). Asterisks (*) mark significant differences between auditory and audiovisual mean SRTs. Pound signs (#) mark significant differences between visual and audiovisual mean SRTs, (student's t test, Bonferroni corrections, p<0.05). Significant differences between unisensory and audiovisual marginal means (ANCOVA, target eccentricity as a covariate) were identical to those reported for simple means (pair wise comparison, Bonferroni corrections, p<0.05). Arrows mark audiovisual conditions in which the mean was significantly shorter (black) or longer (grey) than the shortest unisensory SRT.
Figure 8. Cumulative SRT distributions for audiovisual trials combining different luminous intensities with a quiet sound. Auditory (short dash), visual (long dash), observed audiovisual (solid) and predicted Miller boundary (dot-dash) SRT cumulative distributions are plotted for each visual stimulus (A-F), paired with the -9 dB sound. Insets show the Miller boundary minus the observed audiovisual cumulative probability for each SRT value. Even though unisensory cumulative distributions greatly overlapped, the observed audiovisual distribution was not significantly enhanced beyond the upper boundary of any race-model prediction, (KS-test, Bonferroni correction, p>0.05 data shown: bird ‘J’; p>0.05 data not shown, bird ‘N’).
Figure 9. Cumulative SRT distributions for audiovisual trials combining different SPLs with a bright visual component.

SRT distributions are plotted as cumulative probability plots, and insets show the predicted lower boundary (Grice boundary) of the race model minus the observed cumulative probability for each SRT value. Asterisks (*) mark significant negative violations (KS-test, Bonferroni corrected p<0.05, data shown *bird 'N*). One condition (B) yielded a significant positive violation (#), as compared to the upper boundary of the race model (Miller boundary).
Figure 10. Comparison of errors from auditory, visual, and audiovisual trials. Audiovisual marginal mean errors (± SEM) follow the more accurate unisensory error, but are never more accurate. All conventions follow those of Figure 7, which was an equivalent plot of SRT. Asterisks (*) and pound signs (#) denote conditions where the audiovisual error is significantly different from the auditory or visual error, respectively (t-test, Bonferroni corrected, p<0.05). Data from bird 'N'.
Figure 11. Multivariate analysis of saccadic behavior for one stimulus pair.  
A&B) Azimuthal and elevational components of error are plotted as in Figure 6. Audiovisual error followed visual trends (audiovisual slope and marginal means not significantly different from visual, p>0.05). C) Total error is plotted as in Figure 5. Again, audiovisual error followed visual trends (audiovisual slope and marginal means not significantly different from visual, p>0.05). D) Cumulative distribution comparisons of SRT show that audiovisual SRT follows auditory trends. E) Bivariate plot shows audiovisual response has auditory speed with visual accuracy. Data pooled from both birds "N" and "J".
Figure 12. Summary of mean SRT and error for *birds 'N'* (left) and 'J' (right).

SRT and error values obtained in auditory (grey diamond) and audiovisual (black square) trials are shown normalized to each visual condition (V-norm; grey circle; dashed lines). The same unisensory auditory data are reproduced in each panel for comparison to audiovisual data, and the SPLs are shown above the grey diamonds in panels F and L. In general, the black squares are below the grey diamonds and to the left of V-norm. Thus, multisensory responses were typically initiated faster than visual responses and more accurate than auditory responses across stimulus pairs.