Sensory and Motor Systems

Rapid Audiovisual Integration Guides Predictive Actions

Philipp Kreyenmeier,[1,2,*] Anna Schroeger,[3,4,*] Rouwen Cañal-Bruland,[4] Markus Raab,[5,6] and Miriam Spering[1,2,7,8]

https://doi.org/10.1523/ENEURO.0134-23.2023

[1]Department of Ophthalmology & Visual Sciences, University of British Columbia, Vancouver, British Colombia V5Z 3N9, Canada, [2]Graduate Program in Neuroscience, University of British Columbia, Vancouver, British Colombia V6T 1Z2, Canada, [3]Department of Psychology, Justus Liebig University Giessen, 35390 Giessen, Germany, [4]Department for the Psychology of Human Movement and Sport, Friedrich Schiller University Jena, 07743 Jena, Germany, [5]Department of Performance Psychology, German Sport University Cologne, 50933 Cologne, Germany, [6]School of Applied Sciences, London South Bank University, London SE1 0AA, United Kingdom, [7]Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, British Colombia V6T 1Z3, Canada, and [8]Institute for Computing, Information, and Cognitive Systems, University of British Columbia, Vancouver, British Colombia V6T 1Z4, Canada

Abstract

Natural movements, such as catching a ball or capturing prey, typically involve multiple senses. Yet, laboratory studies on human movements commonly focus solely on vision and ignore sound. Here, we ask how visual and auditory signals are integrated to guide interceptive movements. Human observers tracked the brief launch of a simulated baseball, randomly paired with batting sounds of varying intensities, and made a quick pointing movement at the ball. Movement end points revealed systematic overestimation of target speed when the ball launch was paired with a loud versus a quiet sound, although sound was never informative. This effect was modulated by the availability of visual information; sounds biased interception when the visual presentation duration of the ball was short. Amplitude of the first catch-up saccade, occurring ∼125 ms after target launch, revealed early integration of audiovisual information for trajectory estimation. This sound-induced bias was reversed during later predictive saccades when more visual information was available. Our findings suggest that auditory and visual signals are integrated to guide interception and that this integration process must occur early at a neural site that receives auditory and visual signals within an ultrashort time span.

Key words: eye movements; interception; multisensory integration; perception-action; prediction

Significance Statement

Almost all everyday actions, from catching a ball to driving a car, rely heavily on vision. Although moving objects in our natural visual environment also make sounds, the influence of auditory signals on motor control is commonly ignored. This study investigates the effect of sound on vision-guided interception. We show that sound systematically biases interception movements, indicating that observers associate louder sounds with faster target speeds. Measuring eye movements during interception revealed that vision and sound are integrated rapidly and early in the sensory processing hierarchy. Training and rehabilitation approaches in sports and medicine could harness the finding that interceptive hand movements are driven by multisensory signals and not just vision alone.

Received April 23, 2023; accepted July 22, 2023; First published August 17, 2023.

The authors declare no competing financial interests.

Author contributions: P.K., A.S., and M.S. designed research; P.K. and A.S. performed research; P.K. and A.S. analyzed data; P.K., A.S., R.C.-B., M.R., and M.S. wrote the paper.

Introduction

When intercepting a rapidly moving object with our hands—swatting a fly or catching a ball—we rely heavily on vision. Humans and other animals direct their eyes at moving objects of interest to sample critical visual information, such as the position of the object, speed, and acceleration (Kreyenmeier et al., 2022; Brenner et al., 2023), and to increase performance accuracy (Spering et al., 2011; Diaz et al., 2013; Borghuis and Leonardo, 2015; Michaiel et al., 2020; Fooken et al., 2021). However, other sensory modalities also supply information that might be used to guide behavior in interception tasks. Indeed, in goalball—an interceptive sport for visually impaired athletes—players rely solely on auditory information to locate and intercept a ball (https://goalball.sport/wp-content/uploads/2022/04/IBSA-Goalball-Rules-and-Regulations-2022-2024-v1.1-4-Feb-22.docx-Summary-Change-document.pdf). Our study addresses the question of whether and under which conditions vision-guided interceptive actions rely on sound information in normally sighted observers.

In our natural environment, object motion is almost always accompanied by sound, which can alter visual motion judgements (Sekuler et al., 1997; Soto-Faraco et al., 2003; Senna et al., 2015; Carlile and Leung, 2016; Meyerhoff et al., 2022; Wessels et al., 2022). For instance, hitting a ball with a bat or racket creates an impact sound, and its volume provides information about hit intensity and launch speed. Accordingly, impact sounds can bias perceived ball-bounce locations and perceptual ball speed judgements, suggesting that observers use auditory information to predict ball trajectories (Cañal-Bruland et al., 2018; 2022). When integrating information from different modalities, Bayesian models of multisensory integration predict that sensory signals are combined depending on the uncertainty of the different sensory signals (Ernst and Banks, 2002; Alais and Burr, 2004; Körding et al., 2007; Angelaki et al., 2009). Following this framework, auditory signals may bias visual perception most in tasks with high visual uncertainty such as when viewing conditions are poor (e.g., visual blurring of the target; Schroeger et al., 2021) or visual information is sparse (e.g., short visual presentation durations; Spering et al., 2011).

Our study probes this interaction between visual uncertainty and auditory cues in a real-world-inspired, fast-paced movement interception task during which observers track the brief launch of a simulated baseball moving across the screen and intercept it at a predicted location with a quick pointing movement (Fig. 1A). We manipulated the sound volume of the simulated ball launch and visual uncertainty by varying the visual presentation duration of the ball. At the shortest visual presentation duration, the ball was only visible for 100 ms, a duration that pushes the perceptual system to the limits as it is close to the minimal delay of motion detectors (van de Grind et al., 1986). In this challenging task, we measured observers' eye and hand movements toward the ball as indicators of observers' abilities to estimate ball speed and predict the ball trajectory. We hypothesized, first, that auditory cues would systematically bias ball speed estimation. Specifically, we expected that observers would overestimate speed when the ball launch was accompanied by a loud batting sound (indicating a harder hit and higher launch speed) compared with a quiet batting sound (indicating a softer hit and lower launch speed). Second, we expected that the influence of the auditory cue would scale with visual certainty, implying that observers rely more on the auditory cue when visual presentation durations are short, in line with the assumption that auditory and visual cues are combined by weighing them according to their uncertainty (Fig. 1B). Further, measuring continuous eye movements during this track-intercept task allows us to investigate the time point at which auditory information interacted with visual target speed information and biased observers' estimates of the target trajectory.

Materials and Methods

Participants

We show data from 16 healthy adults (25.5 ± 4.7 years; 11 females, 2 authors). This sample size was determined using an a priori power analysis in G*Power (Faul et al., 2007; power = 0.80; alpha = 0.05) with an effect size of $\eta_p^2 = 0.34$ (main effect of sound volume on estimated speed) derived from pilot data. All observers had normal or corrected-to-normal visual acuity. Study protocols were approved by the University of British Columbia Behavioral Research Ethics Board. Observers were compensated at a rate of $10 CAD per hour.

Apparatus

The experimental setup combined a high-resolution stimulus display with eye and hand tracking. Display and data collection were controlled by a PC (NVIDIA GeForce GTX 1060 graphics card) using MATLAB (version 9.10.0, MathWorks) and the Psychophysics and Eyelink toolboxes (version 3.0.18; Cornelissen et al., 2002; Kleiner et al., 2007). Stimuli were back projected onto a 41.8 × 33.4 cm translucent screen with a PROPixx video projector at a resolution of 1280 × 1024 pixels (120 Hz; VPixx Technologies). Two speakers (S-0264A, Logitech), located 40 cm to the left and right of the screen center, displayed the sound. Observers viewed stimuli binocularly at a distance of 44 cm while their right eye was recorded with an Eyelink 1000 Tower Mount eye tracker (1 kHz; SR Research). The 3D position of each observer's right index finger was recorded with a 3D Guidance trakSTAR (120 Hz; Ascension Technology).

This work was supported by a University of British Columbia International Doctoral Fellowship (P.K.), a German Academic Exchange Service fellowship (A.S.), German Research Foundation grants CA 635/2-2 (R.C-B.) and RA 940/15-2 (M.R.), and a National Sciences and Engineering Research Council of Canada Discovery Grant and Accelerator Supplement (M.S.).

We thank Anna Montagnini and members of the Spering lab for comments.

*P.K. and A.S. share first authorship.

Correspondence should be addressed to Philipp Kreyenmeier at philipp.kreyenmeier@gmail.com.

https://doi.org/10.1523/ENEURO.0134-23.2023

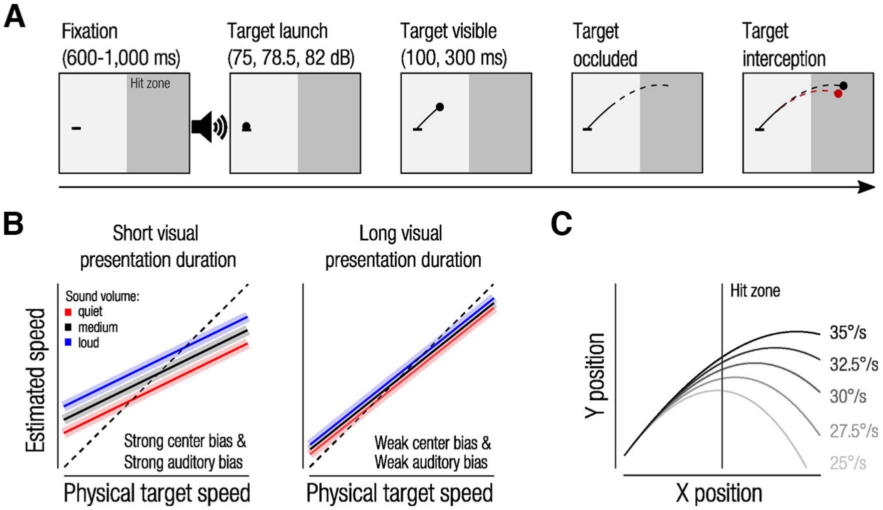

Figure 1. ***A***, Timeline of a single trial. Black lines represent the visible (solid lines) and invisible (dashed lines) parts of the target trajectory. Observers received visual feedback of their finger position (right, red dot) and target position at time of interception (black dot). Red dashed line illustrates the trajectory that best fit the interception position. ***B***, Illustration of hypotheses. Dashed diagonal indicates veridical speed judgments. For short visual presentation durations (high visual uncertainty), we expect a strong regression in estimated speed toward the mean physical target speed (center bias). In addition, we expect that sound volume induces a systematic bias in observers' speed estimates (slower for quiet sounds, faster for loud sounds). Conversely, for long visual presentation durations, we expect less regression toward the mean and only a weak sound-induced bias, indicating that observers relied almost entirely on visual information to estimate target trajectories. ***C***, The five presented ball trajectories defined by different initial launch speeds (gray lines). Vertical line illustrates the border of the hit zone.

Stimuli, experimental procedure, and design

In each trial, we displayed a small black disk that moved along a parabola, simulating the trajectory of a batted baseball affected by gravity, Magnus effect because of the spin of the ball, and aerodynamic drag force (Fooken et al., 2016; Kreyenmeier et al., 2017). The ball was launched at a constant angle of 35° at one of five launch speeds, resulting in five unique trajectories (Fig. 1*C*). All other parameters (e.g., ball mass) to simulate flyball trajectories were the same as in Fooken et al. (2016). The screen was separated into two zones by varying background luminance; the darker right side served as the hit zone in which observers were asked to intercept the ball (Fig. 1*A*). The sound of a baseball hitting a wooden bat was retrieved from a free online sound library (https://freesound.org/people/SocializedArtist45/sounds/266595/; 44.1 kHz) and played at one of three sound volumes (A-weighted sound pressure levels of 75, 78.5, or 82 dBA) for ~50 ms, coinciding with the time of the ball launch.

Each trial began with a random-duration fixation on a line segment that marked the ball-launch position (Fig. 1*A*). After fixation, the ball was launched, paired with a batting sound at one of the three sound intensities (randomly assigned), and moved for either 100 or 300 ms before disappearing from view (Fig. 1*A*, solid black line segment). Observers were instructed to manually intercept the ball anywhere along its extrapolated trajectory (Fig. 1*A*, dashed black line segment) within the hit zone. On interception, a red dot, indicating the interception location of the finger, and a black dot, showing the actual ball position at interception, provided feedback for the observer.

Observers performed nine practice trials (six of these with the entire target trajectory visible) to familiarize themselves with the task. Batting sounds, visual presentation durations, and physical target speeds were pseudorandomly selected for each trial. The experiment consisted of 420 trials in total [14 repetitions for each possible combination of the conditions batting sound × visual presentation duration × physical target speed = 14 × (3 × 2 × 5) = 420], divided into 7 blocks of 60 trials each. Observers took short breaks between blocks.

Eye and hand movement recordings and analyses

Eye and hand movement data were preprocessed off-line. Filtered eye movement traces [second-order Butterworth filtered with 15 Hz (position) and 30 Hz (velocity) cutoff frequencies] were aligned to the target start position. Saccades were detected when five consecutive frames exceeded a fixed velocity criterion of 30°/s. Saccade onsets and offsets were determined as the nearest reversal in the sign of acceleration before eye velocity exceeded the velocity threshold (saccade onset), and the nearest reversal in the sign of acceleration after eye velocity returned below threshold (saccade offset). We inspected all trials manually and excluded trials in which observers blinked or when the eye tracker lost the signal (3.2% of trials across participants).

Hand position data were filtered using a second-order Butterworth filter (15 Hz cutoff) and then upsampled to 1 kHz by linear interpolation. Hand latency was computed as the first sample exceeding 5% of the peak hand velocity in that trial. Hand movement offset was detected when the finger landed within ±0.80 mm of the screen. If no

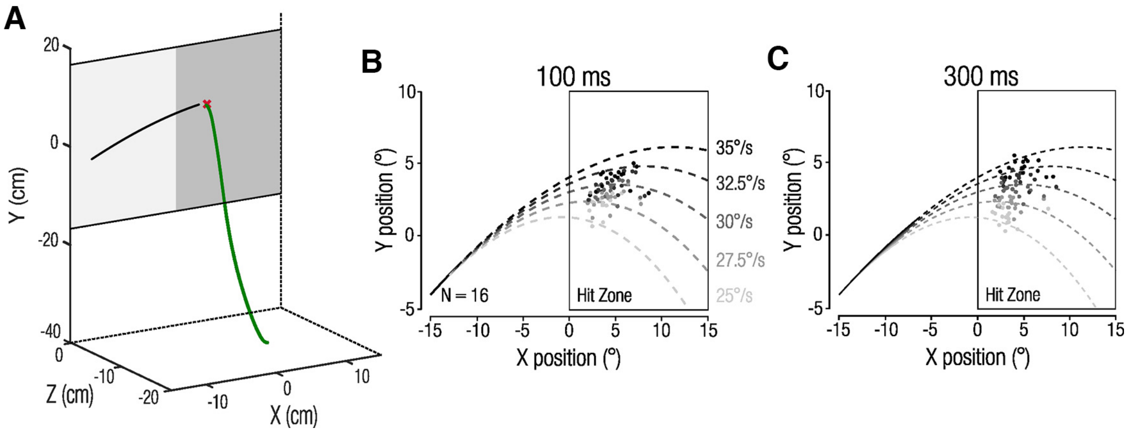

Figure 2. ***A***, Example of a hand position trace (green). Black line represents the 2D target position, and the red cross indicates the interception position. ***B***, ***C***, Mean individual observer 2D interception positions for the 100 ms (***B***) and 300 ms (***C***) visual presentation durations. Each data point indicates one observer's mean interception position per each of the five target speeds.

interception was detected online, interception time and position were determined off-line as the maximum hand position in the *z*-dimension (depth; Fig. 2*A*).

We then used the 2D hand interception position to calculate estimated speed. For each individual trial, we determined which target trajectory best fit the observed interception position (Fig. 1*A*, red dot), as follows. We simulated 600 target trajectories with launch speeds ranging from 0.1 to 60°/s in 0.1°/s steps. We then determined the trajectory (Fig. 1*A*, red dashed line) that produced the smallest Euclidian distance to the interception position. The corresponding target speed that best fit the observed interception position was labeled the estimated speed for that trial. This analysis assumes that observers correctly associate different launch speeds with the different target trajectories (Fig. 1*C*). We confirmed this assumption by analyzing both the vertical and horizontal interception errors, which directly reflect extrapolation errors (de la Malla et al., 2018).

The same analysis was repeated using the eye position at the time of interception to compare how well target speed was estimated based on hand and eye interception. In trials in which observers made a saccade at the time of interception, we used eye position at the offset time for this saccade for the analysis. Next, we analyzed saccade amplitudes during each trial to later obtain a readout of predicted target trajectories at different time points (see below). On average, observers made 2.8 ± 0.7 (mean ± SD) saccades during a trial. We analyzed the amplitude of the first catch-up saccade after target onset as an indicator of early trajectory estimation. After the first catch-up saccade, and after the target disappeared, observers typically made one or two subsequent saccades that brought the eye to the predicted interception location. To account for the varying number of saccades during this later phase of the trial, we calculated the cumulative saccade amplitude (i.e., sum of amplitudes of all subsequent saccades in a trial) as an indicator of late trajectory estimation.

Statistical analyses

To assess effects of sound volume and visual presentation duration on our dependent variables—speed estimates based on interception end points and vertical saccade amplitudes—we first applied a within-subject *z* score outlier detection (data points were excluded if they were >3 SDs from an observer's mean). We then calculated observers' means per condition and fed the data into a repeated measures (rm) ANOVA with an alpha level of 0.05. To correct for multiple comparisons within multiway ANOVA, we applied a sequential Bonferroni correction (Cramer et al., 2016); a Bonferroni correction was also applied to all *post hoc* comparisons (two-sided, paired *t* tests).

In addition to testing these main effects, we also assessed whether physical target speed predicted estimated speed by applying a linear mixed model with physical target speed (continuous predictor), visual presentation duration (categorial predictor), and the interaction term as both fixed and random effects and observers as grouping variable. A linear mixed model was used to obtain regression slopes between physical target speed and speed estimates and to test whether speed estimates scaled more accurately with physical target speed when targets were presented for 300 ms versus 100 ms. All statistical analyses were performed in R software (R Core Team, 2022; www.r-project.org) using RStudio (http://www.rstudio.com/) and the afex (https://CRAN.R-project.org/package=afex), dplyr (https://CRAN.R-project.org/package=dplyr), and ez (http://github.com/mike-lawrence/ez) packages.

Results

Observers tracked the brief launch of the simulated baseball and then intercepted it with a quick pointing movement along the predicted trajectory within a hit zone (Fig. 2*A*). In our task, observers had to rely on visual information of target speed during the brief visual presentation of the ball to extrapolate and intercept it accurately. We predicted that the brief visual presentation durations of 100 or 300 ms would result in conditions of low (short presentation) and high (longer presentation) visual certainty.

Mean 2D interception positions show that observers intercepted targets along their predicted trajectories and discriminated between different target trajectories in both

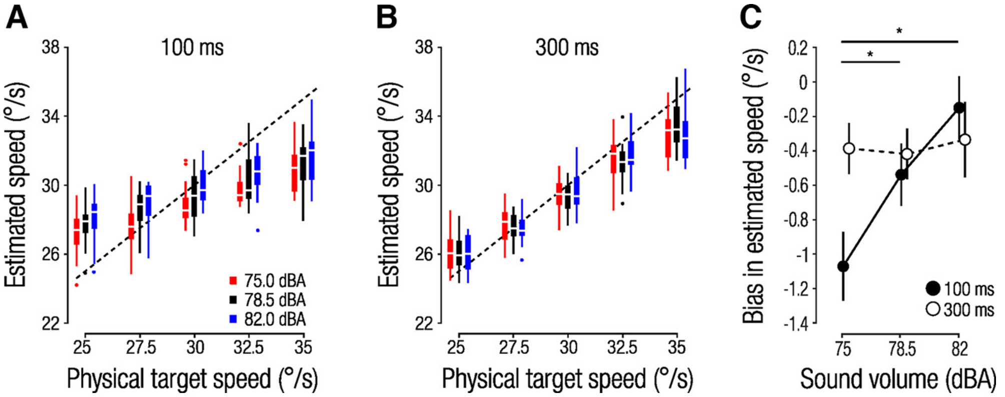

Figure 3. ***A***, ***B***, Box plots of estimated target speed ($n = 16$) as a function of physical target speed. Colors denote sound volume conditions, and dashed lines indicate veridical estimates. ***A***, 100 ms condition; ***B***, 300 ms condition. ***C***, Effect of sound volume on the bias in estimated speed averaged across physical target speeds, separately for the 100 ms (filled circles) and 300 ms (open circles) condition. Circles and error bars denote the mean ± 1 within-subject standard error of the mean (SEM); significant *post hoc* comparisons, $^*p < 0.05$).

the 100 ms (Fig. 2*B*) and 300 ms conditions (Fig. 2*C*). However, interception end points strongly regressed toward the intermediate trajectory in the 100 ms condition, indicating that observers were uncertain about the target trajectory. In contrast, in the 300 ms condition, observers intercepted balls more accurately along their trajectories (Fig. 2*C*).

Auditory cues bias target speed estimates when visual information is uncertain

We predicted that sound volume of the bat-ball contact would systematically bias observers' speed estimates (quiet sounds indicating a softer hit and lower launch speed; loud sounds indicating harder hits and higher speed). Perceptual studies on multisensory cue combination indicate that sensory cues are weighed according to their uncertainty. We thus predicted that under high visual uncertainty (short visual presentation duration), target speed estimates show a systematic sound-induced bias. Under high visual uncertainty, observers are known to rely more strongly on the average speed of all physical targets when judging their trajectories (Jazayeri and Shadlen, 2010; Petzschner et al., 2015). We would therefore expect poor scaling of speed estimates with physical target speed (i.e., a strong center bias) in addition to the systematic sound-induced bias. Conversely, under low visual uncertainty, we expect speed estimates to scale more accurately with physical target speed (i.e., weak center bias) and to be less influenced by the auditory cue (Fig. 1*B*). To test these predictions, we measured observers' speed estimate as the primary outcome measure. Figure 3 shows observers' estimated speed as a function of physical target speed, separately for each sound volume. If speed estimates were accurate, they would fall along the diagonal (dashed line). First, we ran a linear mixed model with physical target speed as a continuous predictor and visual presentation duration as a categorial predictor. Physical target speed was a significant predictor of estimated speed for both visual presentation durations (100 ms, $\beta = 0.37$, $t_{(15)} = 9.66$, $p < 0.001$; 300 ms, $\beta = 0.72$, $t_{(15)} = 14.94$, $p < 0.001$). In line with our predictions, we found a significant difference between slopes for the 100 and 300 ms conditions ($\beta = -0.35$, $t_{(15)} = 15.92$, $p < 0.001$), confirming that observers' speed estimates regressed more toward the mean (indicating high visual uncertainty) in the 100 ms condition compared with the 300 ms condition. Accordingly, the mean 2D interception error was higher in the 100 ms (2.66° ± 0.46°) compared with the 300 ms condition (2.06° ± 0.44°; $t_{(15)} = 10.86$, $p < 0.001$). Together, these findings show that an additional 200 ms of target visibility provide significantly more visual information used to enhance observers' speed estimates.

Next, we asked whether and under which conditions sound volume influenced speed estimates. We hypothesized that sound volume would systematically bias observers' speed estimates and that this bias would depend on the certainty of the visual speed signal. Accordingly, we found that observers systematically underestimated speed when the ball launch was paired with a quiet batting sound and overestimated speed when the ball was paired with a loud batting sound. This effect was consistent across all target speeds at short visual presentation duration (Fig. 3*A*). Conversely, at long visual presentation durations, sound volume did not systematically affect estimated speed (Fig. 3*B*). To assess the differential effects of sound volume at different visual presentation durations we calculated each observer's bias in speed estimation across physical target speeds (mean difference between estimated speed and physical target speed; Fig. 3*C*). A 2 (visual presentation duration) × 3 (sound volume) rmANOVA revealed a significant main effect of sound volume ($F_{(2,30)} = 4.91$, $p = 0.029$, $\eta_p^2 = 0.25$) and no main effect of visual presentation duration ($F_{(1,15)} = 0.60$, $p = 0.45$, $\eta_p^2 = 0.04$). A significant sound volume × visual presentation duration interaction ($F_{(1.43,21.46)} = 20.30$, $p < 0.001$, $\eta_p^2 = 0.58$) confirmed the profound effect of auditory cues on manual interception when visual information is sparse but not when the target is presented sufficiently long to base speed estimation for interception on visual information alone. These findings show that when visual information was sparse and

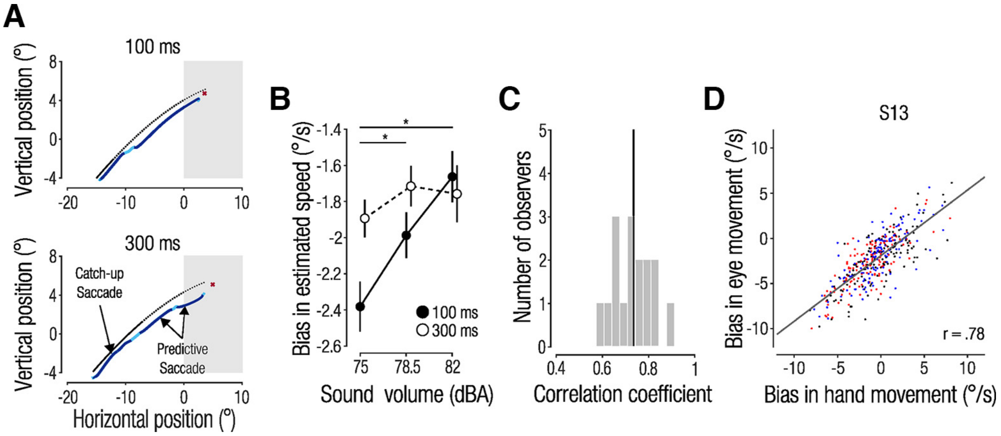

Figure 4. ***A***, Two-dimensional eye position traces of two representative trials. Bright blue segments indicate smooth pursuit, continuous tracking of moving targets with the eyes, and dark blue segments indicate saccades. Solid and dashed black lines represent the visible and invisible portions of the target trajectory. The shaded area represents the hit zone. ***B***, Effect of sound volume on estimated speed based on final eye position. ***C***, Histogram of trial-by-trial correlation coefficients from all observers. Black line indicates mean across observers. ***D***, Trial-by-trial correlation of one representative observer.

thus uncertain, speed estimates were strongly biased toward the mean and systematically influenced by the auditory cue. Conversely, when visual uncertainty was low, estimated speed scaled almost perfectly with physical target speed (weak center bias) and showed no impact of auditory cues.

Eye movements reveal temporal dynamics of audiovisual integration

The extent to which observers relied on sound depended on the certainty of the visual speed signal, that is, visual presentation duration (low certainty for short, high certainty for longer presentations). The impact of the auditory signal decreased with increasing visual presentation duration. To assess how differences in auditory signal use in the long and short visual presentation duration conditions unfolded over time, we analyzed observers' continuous eye movements during the interception task.

Observers tracked the simulated baseball with their eyes using a combination of smooth pursuit and saccadic eye movements (Fig. 4*A*). They typically made an early catch-up saccade shortly after target onset (mean = 125, SD = 38 ms). Subsequent predictive saccades were made after target disappearance to the predicted interception location. Eye movement endpoints, based on the 2D eye position at the time of interception, reflect observers' speed estimates. Figure 4*B* shows that observers underestimate speed in the presence of a quiet sound and overestimate speed when paired with a louder sound, akin to observations for manual interception responses (Fig. 3*C*). Accordingly, speed estimates based on eye and hand movement end points were strongly correlated on a trial-by-trial basis with a mean correlation of $r = 0.73$ (measured across physical target speeds and sound volumes; Fig. 4*C*; trial-by-trial correlation of one representative observer depicted in Fig. 4*D*).

We next assessed whether eye movements can indicate the time point at which the auditory cue first started influencing observers' trajectory estimates. Specifically, we asked whether the first catch-up saccade made after target onset (initiated with a mean latency of 125 ms) was already influenced by sound volume. This would indicate early audiovisual integration. By contrast, an effect only on subsequent predictive saccades, made later in the trial, would indicate that integration processes take longer. We analyzed the amplitude of the first catch-up saccade and the combined amplitudes of subsequent, predictive saccades. If sound volume biases saccades similarly to what we observed for eye and hand interception end points, we would expect that loud sounds lead to larger saccades (following a trajectory with higher launch speed), and that quiet sounds lead to smaller saccade amplitudes. For these analyses, we excluded trials where the first catch-up saccade was made in anticipation of target onset ($\leq$50 ms latency, 3.9% of trials).

Figure 5*A* shows the horizontal amplitude of the first catch-up saccade plotted against the vertical amplitude, separately for each physical target speed and for the two visual presentation durations. We found an influence of sound volume on the amplitude of the first catch-up saccade, consistently observed across physical target speeds and visual presentation durations. Sound volume exhibited the strongest influence on the vertical saccade amplitude, in line with our observation that interception end points differentiated between trajectories primarily along the vertical axis (Fig. 2*B*,*C*). Feeding the mean vertical saccade amplitudes (averaged across physical target speeds) into a 2 (visual presentation duration) × 3 (sound volume) rmANOVA revealed a main effect of sound volume ($F_{(2,30)} = 26.24$, $p < 0.001$, $\eta_p^2 = 0.64$; Fig. 5*B*). Neither the main effect of visual presentation duration nor the interaction term were significant (all p values > 0.388), indicating that the auditory cue influenced speed estimates early during the trial and before any differences in presentation duration could have had an impact on these estimates. Note that the first catch-up saccade not only showed consistent and similar effects of sound

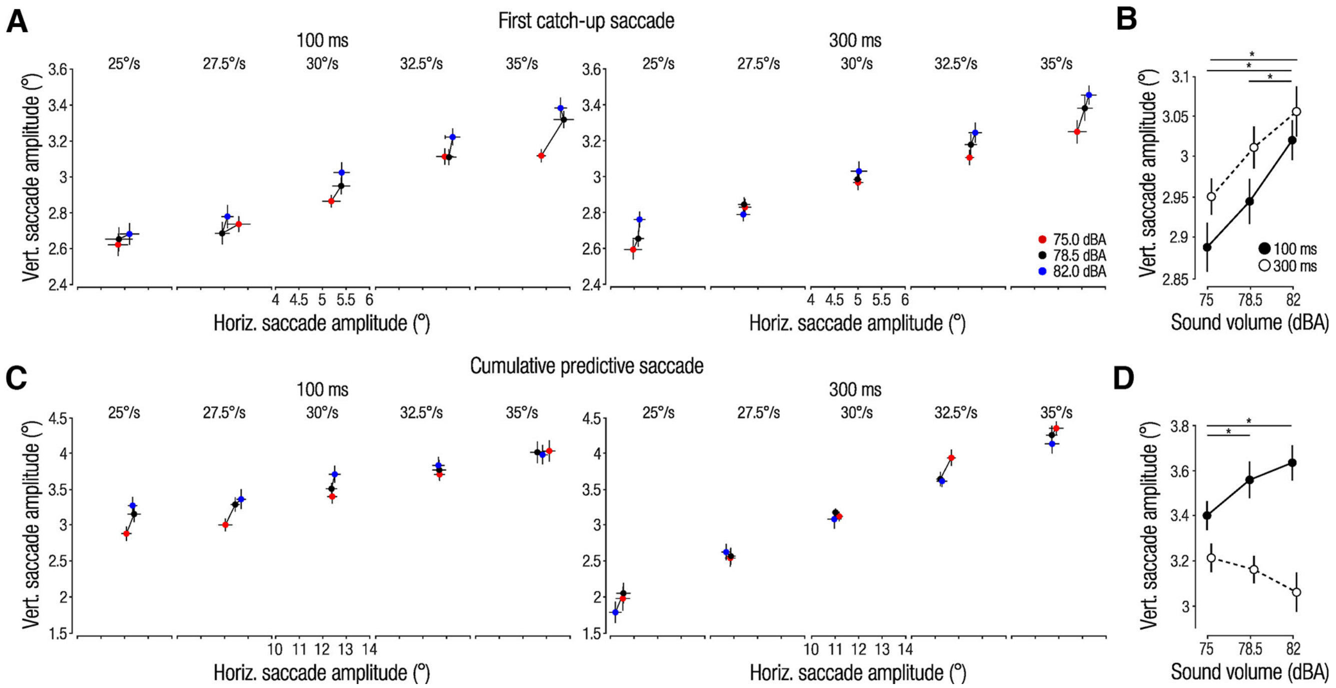

Figure 5. Saccade analyses. ***A***, Effect of sound volume on horizontal and vertical saccade amplitudes for the first catch-up saccade after target onset. ***B***, Vertical saccade amplitudes averaged across physical target speeds. ***C***, ***D***, Cumulative saccade amplitudes of all subsequent saccades. Circles and error bars show means ± 1 within-subject SEM; significant *post hoc* comparisons, *$p < 0.05$. Note different scales between top and bottom panels.

volume between presentation durations but also scaled similarly with physical target speed (Fig. 5*A*). This further indicates that early catch-up saccades were finely tuned to the sensory properties of the target and were programmed before differences between presentation durations emerged.

This early auditory bias, observed for both visual presentation durations, contrasts with our finding that speed estimates based on eye and hand movement end points were only biased for the short but not the long duration. We would therefore expect that subsequent, predictive saccades reverse the early auditory influence when more visual information is available (i.e., in the 300 ms condition). Thus, we next analyzed the combined amplitudes of all subsequent saccades. We used the cumulative saccade amplitude because the number of saccades differed between trials and observers, meaning that a reversal of the early auditory influence could either occur by making smaller or fewer saccades. In line with our expectation, predictive saccades that occurred later in the trial had larger amplitudes with increasing sound volume in the 100 ms condition but smaller amplitudes with increasing sound volume in the 300 ms condition (Fig. 5*C*). Again, sound primarily affected the vertical component of the cumulative saccade amplitudes.

We averaged vertical cumulative saccade amplitudes across physical target speeds (Fig. 5*D*) and fed the data into a 2 (visual presentation duration) × 3 (sound volume) rmANOVA. In line with our expectation of a differential impact of sound volume, depending on availability of visual information, we did not find a main effect of sound volume ($F_{(1.38,20.68)} = 0.71$, $p = 0.454$, $\eta_p^2 = 0.04$) but instead a strong visual presentation duration × sound volume interaction ($F_{(2,30)} = 20.46$, $p < 0.001$, $\eta_p^2 = 0.58$). A significant main effect of visual presentation duration ($F_{(1,15)} = 10.98$, $p = 0.009$, $\eta_p^2 = 0.42$) is likely because of smaller saccades in the 300 ms condition, which generally elicits stronger pursuit. The differential impact of sound volume for the 100 and 300 ms conditions indicates a reversal of the early auditory influence with the availability of additional visual information. This observation was further supported by the finding that predictive saccades in the 300 ms condition scaled more with physical target speed than predictive saccades in the 100 ms condition (Fig. 5*C*).

Discussion

Predicting the trajectory of a moving object is a fundamental ability that allows us to accurately hit, catch, or otherwise intercept targets (Fiehler et al., 2019). Most research on interception focuses solely on vision to form trajectory predictions and guide interceptive hand movements (Brenner and Smeets, 2018; Fooken et al., 2021). Yet, in our natural environment, object motion is typically accompanied by sounds that can provide additional information about the motion of an object. Here, we show that auditory signals are used in combination with visual motion information to estimate target speed for interceptive actions. Using a rapid track-intercept task in which a visual trajectory was paired with batting sounds of varying intensities we present three key findings. (1) Sound

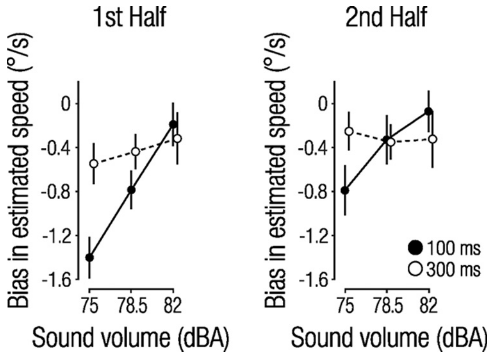

Figure 6. Bias in estimated speed split separately for the first and second half of the experiment. Solid lines and filled circles represent 100 ms, and dashed lines and open circles represent the 300 ms condition. Circles and error bars show mean ± 1 within-subject SEM.

volume of bat-ball contact systematically influences interception responses, extending well-known effects of audiovisual integration on perception to interceptive actions. (2) Integration of auditory cues and visual information depends on the certainty of the visual signal; auditory cues influence speed estimates only when visual information is sparse. (3) Audiovisual integration occurred as early as the first catch-up saccade (initiated 125 ms after target onset on average); with the availability of additional visual information, the early sound bias was reversed. The temporal dynamics of audiovisual integration was revealed by analyzing continuous eye movements during this task. In our experiment, sound volume was never informative of physical target speed, precluding the possibility that our results were solely caused by learning to associate certain sound volumes to certain target trajectories. Instead, our findings likely reflect a natural association between sound volume and relative target speed gained through lifelong experience. Under similar environmental conditions, particularly when the target is always at the same distance from the observer, louder sounds will typically correspond to higher target speeds. When splitting our data between first and second halves of the experiment, we found that the auditory influence was stronger during the first half of the experiment (Fig. 6). This indicates a strong association between sound volume and target speed that decreased with increasing task experience. Together, these findings highlight the important contribution of auditory cues for vision-guided actions, particularly in situations where visual information is sparse or uncertain. These results build on a long line of literature on audiovisual signal integration for perceptual tasks (Ernst and Bülthoff, 2004). The novelty of our findings lies in uncovering how auditory information contributes to vision-guided interception, a fundamental ability for everyday interactions.

By manipulating the visual presentation duration of the target, we revealed that the use of auditory cues critically depends on the uncertainty of the visual motion signal. This finding is aligned with previous perceptual studies on multisensory cue combination that used Bayesian observer models and show that prior information and sensory evidence are combined depending on their respective uncertainty (Ernst and Banks, 2002; Alais and Burr, 2004; Körding et al., 2007; Angelaki et al., 2009). Congruently, we found that speed estimates were only influenced by auditory cues when visual information was sparse, whereas the auditory cue was largely ignored when sufficient visual information was provided. Moreover, we observed a strong center bias in speed estimates when visual information was uncertain. This type of finding is typically interpreted to indicate use of a prior based on the statistics of the stimuli used (Jazayeri and Shadlen, 2010; Petzschner et al., 2015; Chang and Jazayeri, 2018). Alternatively, priors can also be derived from statistics of our natural environment. Studies on visual (Stocker and Simoncelli, 2006) and auditory motion perception (Senna et al., 2015) revealed that observers typically rely on a slow-motion prior. Our finding that observers generally undershot target trajectories (Fig. 3) fits with those results.

It is important to note that any variation in ball presentation duration might not only affect visual uncertainty but might also impact the reliability of the auditory cue. The auditory cue was always presented at the time of ball launch, whereas visual information was either presented for 100 or 300 ms. Therefore, a longer visual presentation might potentially downweigh the reliability of the auditory cue as more visual information was provided after the sound. Because we did not independently manipulate the reliability of both cues, we cannot rule out that the reliability of the auditory cue might have had an impact on our results.

Whereas our approach did not allow us to fully test Bayesian cue integration, future studies could include unimodal (auditory and visual) conditions in addition to the audiovisual condition to directly test predictions of Bayesian cue combination in the context of interception. Moreover, including an auditory-only condition could allow assessment of whether observers naturally associate auditory intensities of bat-ball contact with ball launch speed even in the absence of visual information.

Our interception task was inspired by baseball. We used a visual target that moved along a simulated batted baseball and a naturalistic batting sound. Based on real-world Major League Baseball data, it was recently shown that baseball batters rely on prior knowledge and visual cues, for example, a pitcher's posture and hand position when estimating where to swing (Brantley and Körding, 2022). Simple cues and heuristics are critical in baseball, where hitters only have a few hundred milliseconds to decide whether and where to swing (Gray and Cañal-Bruland, 2018). In this or similar rapid decision-making contexts, auditory cues may provide a critical advantage because combining them with visual cues can reduce uncertainty (Alais and Burr, 2004). Yet, future studies are needed to assess whether athletes rely on auditory cues of bat-ball contacts, in addition to prior knowledge and visual signals during real-world interceptive sports, as our findings suggest.

Eye movements as sensitive indicators of audiovisual integration

Eye movements are a natural, instinctive behavior in tasks that require fine-motor interactions with a visual object. When manually intercepting, hitting, or catching an object, observers track its trajectory until the point of interception (Mrotek and Soechting, 2007; Fooken et al., 2021). The continuous nature of these movements provides an opportunity to relate their kinematics to ongoing cognitive task processes, such as decision-making (Spering, 2022). Here, we used observers' continuous eye movements to probe the temporal dynamics of audiovisual integration. We observed a systematic influence of the auditory cue on the first catch-up saccade, which was initiated, on average, 125 ms after target onset. At this early time point, louder sound volumes evoked larger saccade amplitudes. If additional visual information was available (long visual presentation duration), subsequent saccades reversed this early auditory effect. This finding suggests that the integration of auditory and visual signals can occur at a very short timescale, in line with findings showing early effects of audiovisual cues on pupil dilation and simple saccadic decision-making (Wang et al., 2017). Previous studies have identified the superior colliculus—a midbrain structure that is also involved in the control of eye movements (Sparks, 1999)—as a key hub of audiovisual integration (Stein and Stanford, 2008). Visual and auditory signals reach this brain structure within 80 ms (Ito et al., 2021), making this area an excellent candidate for short-latency audiovisual integration. In parallel, visual and auditory signals could also be integrated in cortical sensory areas such as the middle temporal cortex, an area traditionally dedicated to early visual motion processing (Rezk et al., 2020).

We conclude that auditory signals significantly and systematically have an impact on vision-guided interceptive actions. This influence was strongest when visual information was sparse. We show that noninvasive, time-sensitive eye movement measurements can provide new behavioral evidence for early and rapid integration of auditory and visual signals.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262.

Angelaki DE, Gu Y, DeAngelis GC (2009) Multisensory integration: psychophysics, neurophysiology, and computation. Curr Opin Neurobiol 19:452–458.

Borghuis BG, Leonardo A (2015) The role of motion extrapolation in amphibian prey capture. J Neurosci 35:15430–15441.

Brantley JA, Körding KP (2022) Bayesball: Bayesian integration in professional baseball batters. bioRxiv 511934. https://doi.org/10.1101/2022.10.12.511934.

Brenner E, Smeets JBJ (2018) Continuously updating one's predictions underlies successful interception. J Neurophysiol 120:3257–3274.

Brenner E, de la Malla C, Smeets JBJ (2023) Tapping on a target: dealing with uncertainty about its position and motion. Exp Brain Res 241:81–104.

Cañal-Bruland R, Müller F, Lach B, Spence C (2018) Auditory contributions to visual anticipation in tennis. Psychol Sport Exerc 36:100–103.

Cañal-Bruland R, Meyerhoff HS, Müller F (2022) Context modulates the impact of auditory information on visual anticipation. Cogn Res Princ Implic 7:76.

Carlile S, Leung J (2016) The perception of auditory motion. Trends Hear 20:2331216516644254.

Chang CJ, Jazayeri M (2018) Integration of speed and time for estimating time to contact. Proc Natl Acad Sci USA 115:E2879–E2887.

Cornelissen FW, Peters EM, Palmer J (2002) The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. Behav Res Methods Instrum Comput 34:613–617.

Cramer AOJ, van Ravenzwaaij D, Matzke D, Steingroever J, Wetzels R, Grasman RPPP, Waldorp LJ, Wagenmakers JE (2016) Hidden multiplicity in exploratory multiway ANOVA: prevalence and remedies. Psychon Bull Rev 23:640–647.

De la Malla C, Smeets JBJ, Brenner E (2018) Errors in interception can be predicted from errors in perception. Cortex 98:49–59.

Diaz GJ, Cooper J, Rothkopf C, Hayhoe M (2013) Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. J Vis 13(1):20, 1–14.

Ernst M, Banks M (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8:162–169.

Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191.

Fiehler K, Brenner E, Spering M (2019) Prediction in goal-directed action. J Vis 19(9), 10, 1–21.

Fooken J, Yeo SH, Pai DK, Spering M (2016) Eye movement accuracy determines natural interception strategies. J Vis 16(14):1, 1–15.

Fooken J, Kreyenmeier P, Spering M (2021) The role of eye movements in manual interception: a mini-review. Vision Res 183:81–90.

Gray R, Cañal-Bruland R (2018) Integrating visual trajectory and probabilistic information in baseball batting. Psychol Sport Exerc 36:123–131.

Ito S, Si Y, Litke AM, Feldheim DA (2021) Nonlinear visuoauditory integration in the mouse superior colliculus. PLoS Comput Biol 17: e1009181.

Jazayeri M, Shadlen MN (2010) Temporal context calibrates interval timing. Nat Neurosci 13:1020–1026.

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. Perception 36:1–16.

Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JR, Shams L (2007) Causal inference in multisensory perception. PLoS One 2: e943.

Kreyenmeier P, Fooken J, Spering M (2017) Context effects on smooth pursuit and manual interception of a disappearing target. J Neurophysiol 118:404–415.

Kreyenmeier P, Kämmer L, Fooken J, Spering M (2022) Humans can track but fail to predict accelerating objects. eNeuro 9:ENEURO.0185-22.2022–15.

Meyerhoff HS, Gehrer NA, Merz S, Frings C (2022) The beep-speed illusion: non-spatial tones increase perceived speed of visual objects in a forced-choice paradigm. Cognition 219:104978.

Michaiel AM, Abe ETT, Niell CM (2020) Dynamics of gaze control during prey capture in freely moving mice. Elife 9:e57458.

Mrotek LA, Soechting JF (2007) Target interception: hand-eye coordination and strategies. J Neurosci 27:7297–7309.

Petzschner FH, Glasauer S, Stephan KE (2015) A Bayesian perspective on magnitude estimation. Trends Cogn Sci 19:285–293.

Rezk M, Cattoir S, Battal C, Occelli V, Mattioni S, Collignon O (2020) Shared representation of visual and auditory motion directions in the human middle-temporal cortex. Curr Biol 30:2289–2299.e8.

Schroeger A, Tolentino-Castro JW, Raab M, Cañal-Bruland R (2021) Effects of visual blur and contrast on spatial and temporal precision in manual interception. Exp Brain Res 239:3343–3358.

Sekuler R, Sekuler A, Lau R (1997) Sound alters visual motion perception. Nature 385:308.

Senna I, Parise CV, Ernst MO (2015) Hearing in slow-motion: humans underestimate the speed of moving sounds. Sci Rep 5:14054.

Soto-Faraco S, Kingstone A, Spence C (2003) Multisensory contributions to the perception of motion. Neuropsychologia 41:1847–1862.

Sparks DL (1999) Conceptual issues related to the role of the superior colliculus in the control of gaze. Curr Opin Neurobiol 9:698–707.

Spering M (2022) Eye movements as a window into decision-making. Annu Rev Vis Sci 8:427–448.

Spering M, Schütz AC, Braun DI, and Gegenfurtner KR (2011) Keep your eyes on the ball: smooth pursuit eye movements enhance prediction of visual motion. J Neurophysiol 105:1756–1767.

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:255–266.

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585.

Van de Grind WA, Koenderink JJ, van Doorn AJ (1986) The distribution of human motion detector properties in the monocular visual field. Vision Res 26:797–810.

Wang CA, Blohm G, Huang J, Boehnke SE, Munoz DP (2017) Multisensory integration in orienting behavior: pupil size, microsaccades, and saccades. Biol Psychol 129:36–44.

Wessels M, Zähme C, Oberfeld D (2022) Auditory information improves time-to-collision estimation for accelerating vehicles. Curr Psychol https://doi.org/10.1007/s12144-022-03375-6.