

RESEARCH REPORT

Auditory Rhythm Biases Visual Speed Perception and Shows Tempo-Dependent Signatures During Motion Extrapolation

Yantao Liu¹ | Lihan Chen^{1,2,3,4} 

¹School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China | ²National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing, China | ³Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China | ⁴Beijing Key Laboratory of Brain-Computer Interface and Mental Health Modulation, Beijing, China

Correspondence: Lihan Chen (clh@pku.edu.cn)

Received: 23 August 2026 | **Revised:** 8 October 2026 | **Accepted:** 14 October 2026

Associate Editor: Yoland Smith

Keywords: acceleration perception | audiovisual integration | auditory pitch | auditory tempo | eye movements | time-to-contact | visual motion perception

ABSTRACT

Auditory signals can bias visual motion judgements, but how dynamic auditory structure affects variable-speed motion during perception and subsequent extrapolation remains unclear. Across three experiments, we paired changing visual motion with changes in auditory tempo or pitch. In Experiment 1, slow and fast tempi produced robust, opposite shifts in the point of subjective equality for visual speed-change judgements, without reliable evidence for a change in discrimination precision in this paradigm. Experiment 2 revealed a weaker asymmetric pitch effect: Descending pitch shifted judgements toward deceleration, whereas ascending pitch produced no reliable opposite shift. A pooled comparison localized the tempo-pitch difference primarily to the speeding-direction bias rather than psychometric slope. In Experiment 3, the omnibus three-level auditory effect on manual time-to-contact (TTC) was not significant; however, within the two sound-present conditions, accelerating tempo shortened mean manual TTC by approximately 32 ms relative to decelerating tempo. The corresponding primary eye-TTC contrast was small and not significant, although time-resolved eye analyses revealed a transient postocclusion eye-velocity effect together with broader sound-related pupil responses. Leave-one-condition-out model comparison nevertheless showed that the same tempo-by-physical-TTC functional model provided the best held-out prediction for both manual and eye-based estimates. Together, these findings indicate robust auditory-tempo biases in online visual speed judgements and more qualified, tempo-dependent signatures during forward extrapolation after visual information disappears; the absence of a JND effect here should not be generalized to all audiovisual motion tasks.

1 | Introduction

Auditory signals can alter visual motion perception even when they provide no spatial information about the visual target. Moving sounds and temporally structured sounds can bias

apparent motion and perceived visual speed (Kafaligonul and Stoner 2010; Hidaka et al. 2011; Kaya and Kafaligonul 2019; Meyerhoff et al. 2022; for the converse visual influence on perceived auditory speed, see López-Moliner and Soto-Faraco 2007), and auditory temporal rate can influence visual temporal-rate

Abbreviations: AIC, Akaike information criterion; BCI, Bayesian causal inference; BIC, Bayesian information criterion; CI, confidence interval; GLMM, generalized linear mixed model; JND, just noticeable difference; LOCO, leave-one-condition-out; LOCO-logLik, leave-one-condition-out log-likelihood; PSE, point of subjective equality; RMSE, root mean square error; SD, standard deviation; TTC, time-to-contact.

© 2026 Federation of European Neuroscience Societies and John Wiley & Sons Ltd.

All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

judgements more generally (Welch et al. 1986; Recanzone 2003). These findings place audiovisual motion effects within a broader class of crossmodal temporal interactions rather than treating them as simple response facilitation.

A useful theoretical framework is Bayesian causal inference and reliability-weighted multisensory integration (Ernst and Banks 2002). On these accounts, crossmodal cues are combined to the extent that they are inferred to arise from a common event, and their influence depends on relative reliability (Körding et al. 2007; Shams and Beierholm 2010; Shams 2012). This framework does not require every auditory feature to receive the same weight. Visual acceleration estimates are comparatively imprecise (Watamaniuk and Heinen 2003; Brenner et al. 2016; Kreyenmeier et al. 2022; but see Brenner et al. 2014, for precise interceptive timing with gravitationally accelerating targets), whereas audition has high temporal resolution; therefore, global auditory timing may exert greater influence than a less directly temporal feature correspondence. Relatedly, Chen et al. (2018) showed that auditory and visual temporal information can be partially integrated through temporal averaging, providing a concrete precedent for graded crossmodal weighting rather than all-or-none fusion.

The distinction between local temporal anchors and global sequence structure is especially relevant to changing tempo. Listeners are sensitive to gradual tempo changes and use them to anticipate upcoming events (Pouliot and Grondin 2005; Rankin et al. 2009; Cope et al. 2012; van der Steen et al. 2015; Graber and Fujioka 2019). Recent evidence indicates that perceived visual speed can depend on the overall auditory temporal interval structure rather than solely on individual onset or offset anchors (Özkan et al. 2026). If so, a gradual change in auditory tempo could act as a continuously evolving temporal context for visual motion rather than as a sequence of isolated transients. This possibility motivates testing abrupt and gradual audiovisual changes separately.

In the present study, we implemented three experiments. Experiments 1 and 2 were designed to compare a directly temporal auditory cue with a featural correspondence. Experiment 1 manipulated auditory tempo while participants judged abrupt or gradual changes in visual speed. We predicted opposite PSE shifts for slow and fast tempo if auditory rate is mapped onto visual speed. Experiment 2 held event rate constant while pitch ascended or descended. Because pitch-motion correspondences are well documented but need not be symmetric (Maeda et al. 2004; Sadaghiani et al. 2009; King and Prime 2019), we predicted a pitch-dependent bias but did not assume equal and opposite effects for ascending and descending trajectories. Comparing the two experiments therefore tests whether global temporal structure and feature correspondence receive different effective weights.

Auditory information also affects extrapolation after a target becomes invisible (Teramoto et al. 2010; for review, see Hubbard 2015; Battaglini and Ghiani 2021). In prediction-motion and time-to-contact (TTC) tasks, nonspatial auditory features can shorten or prolong judged TTC (Battaglini et al. 2013; Battaglini and Ghiani 2021; Chotsrisuparat

et al. 2017; DeLucia et al. 2016; Patrick and Anderson 2022; Qin et al. 2022; Wessels et al. 2022, 2023). Two broad mechanisms are plausible. Auditory structure could alter the weighting or extrapolation of the stored visual motion history, or it could alter the temporal scale on which an internal prediction unfolds. The latter possibility includes clock-rate accounts in which faster auditory rhythms accelerate an internal timing process, leading to earlier produced responses (Treisman et al. 1990; Penton-Voak et al. 1996; Wearden et al. 1999). Experiment 3 was designed to distinguish at least the functional signatures of these possibilities by presenting changing tempo only after occlusion and by comparing additive, kinematic, and tempo-by-physical-TTC models. We expected auditory tempo to shift TTC in opposite directions and predicted that any direction-specific oculomotor effect should emerge after auditory onset rather than before occlusion; pupil effects were treated as a broader index of sound-related physiological/task state rather than a directional readout of motion prediction. Across the three experiments, our central question is therefore how dynamic auditory structure changes the bias and temporal evolution of visual motion judgements, whether that influence depends on auditory feature type, and whether a common functional mapping can describe explicit and oculomotor extrapolation after visual information disappears.

2 | Materials and Methods

2.1 | Experiment 1: Effect of Sound Tempo on Visual Speed Perception

2.1.1 | Participants

A total of 18 students from Peking University were recruited to participate in the experiment. Data from four participants were excluded from the final analysis: one due to accuracy rates falling below chance levels, one due to near-perfect accuracy (indicating a ceiling effect), and two due to incomplete experimental sessions. The final sample consisted of 14 valid participants (seven males and seven females) with a mean age of 22.07 years ($SD = 2.37$). All participants had normal or corrected-to-normal vision and self-reported normal hearing. Prior to the experiment, all participants provided written informed consent. The study protocol was approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences, Peking University. Sample size was based on comparable within-subject audiovisual speed-perception studies. A sensitivity power analysis (G*Power; Faul et al. 2007) indicated that the final sample ($N = 14$) provides 80% power to detect within-subject effects of $d_z \geq 0.81$ (two-tailed paired-samples t -test, $\alpha = 0.05$; Lakens 2013); the observed tempo-induced PSE effects ($|d_z| = 1.03\text{--}1.91$) exceeded this threshold. No participant took part in both Experiment 1 and Experiment 2 (see Participants, Experiment 2).

2.1.2 | Experimental Design

The full experiment included three visual-change modes: abrupt change, gradual change, and a 500-ms gap control between two

motion segments. The abrupt and gradual modes each crossed three auditory conditions (slow tempo, fast tempo, and no sound) and constitute the primary 2×3 analysis reported below. The gap control was accompanied by the two sound-present sequences and served as a first-order speed-discrimination comparison, requiring participants to judge whether postgap speed was faster or slower than pregap speed. Thus, the gap trials were part of the broader experimental design but were not included in the primary abrupt/gradual psychometric analysis.

The method of constant stimuli was used to sample the psychometric function. Signed visual speed change was treated as a continuous predictor. PSE was defined as the signed speed-change value yielding 50% “acceleration” responses. Discrimination precision was summarized as JND_{75-50} , the distance from the PSE to the 75% point of the fitted logistic psychometric function; for a logistic slope b , $JND_{75-50} = \ln(3)/|b|$. Positive PSE values indicate a bias toward perceived visual deceleration, whereas negative values indicate a bias toward perceived visual acceleration.

2.1.3 | Stimuli and Apparatus

Stimuli were generated using MATLAB R2021b (MathWorks, Natick, MA) and Psychtoolbox (Brainard 1997; Kleiner et al. 2007; Pelli 1997) and displayed on a 27-in LCD monitor (1920×1080 pixels, 60 Hz). Viewing distance was 58 cm. Auditory stimuli were delivered through Sony MDR-XB550 headphones at 60 dB.

The visual stimulus was a drifting Gabor patch (Gaussian envelope, $\sigma = 1^\circ$; spatial frequency = 2 cycles/degree; contrast = 50%) presented at fixation (Figure 1A). Motion direction was randomly selected from four diagonal directions with equal probability. The initial speed was $6^\circ/\text{s}$. Ten signed speed-change levels were tested ($\pm 6\%$, $\pm 12\%$, $\pm 20\%$, $\pm 30\%$, and $\pm 50\%$), yielding final speeds from 4 to $9^\circ/\text{s}$. Percentage changes were defined using the lower speed as the denominator (e.g., $+50\%$: $6 \times 1.5 = 9^\circ/\text{s}$; -50% : $6 \div 1.5 = 4^\circ/\text{s}$). In gradual trials, visual speed changed linearly rather than stepping at the midpoint: Each gradual profile was

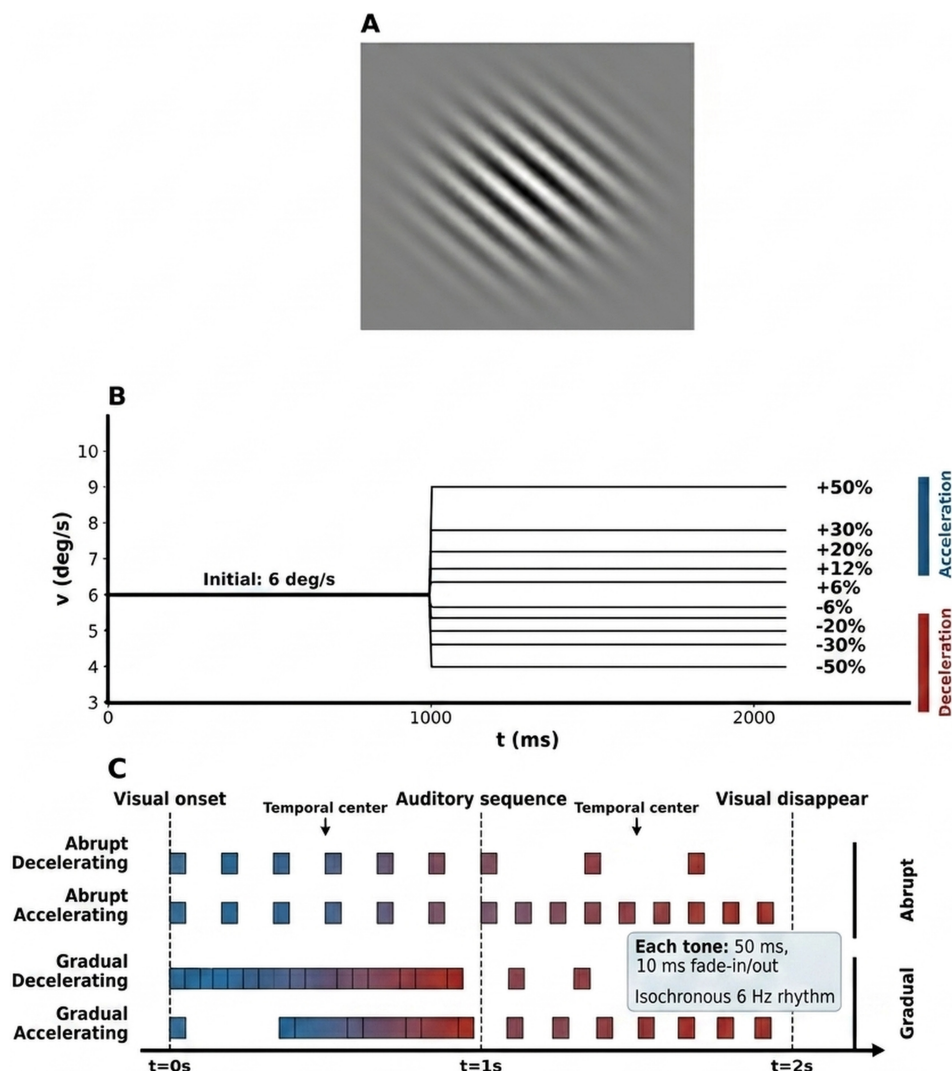


FIGURE 1 | Stimuli and setup for Experiment 1. (A) Example of the moving Gabor stimulus. (B) Visual-speed profiles for the 10 signed speed-change levels in the abrupt condition; the corresponding gradual condition changes continuously rather than stepping at the midpoint. (C) Temporal relationship between the visual motion and auditory sequences. The first 1000 ms establish the initial visual speed and 6-Hz auditory rate. In abrupt trials, visual speed and auditory tempo transition around the midpoint; in gradual trials, the changes evolve continuously during the second half of the sequence. Slow tempo decreases the auditory event rate and fast tempo increases it. The no sound control contains no auditory sequence.

constructed from the corresponding abrupt profile by connecting the midpoints of the prechange and postchange segments, so that speed remained at the initial value for 0–500 ms, changed linearly over 500–1500 ms (crossing the abrupt step value at the 1000-ms midpoint), and remained at the final value for 1500–2000 ms (see schematic Figure 2).

Auditory stimuli were 50-ms, 500-Hz sinusoidal tones with 10-ms onset and offset ramps. The initial presentation rate was 6 Hz. In the fast tempo condition, rate increased toward 9 Hz; in the slow tempo condition, rate decreased toward 3 Hz. In abrupt trials, the rate transition occurred at the midpoint of the 2-s visual sequence; in gradual trials, auditory rate changed linearly during the second half of the sequence. Tone onsets were centered within their nominal temporal periods rather than locked to display frames. In gradual trials, the auditory rate changed continuously across the entire 2000-ms sequence: The fast tempo sequence comprised 15 tones with interonset intervals decreasing from 207 to 97 ms, and the slow tempo sequence comprised 10 tones with interonset intervals increasing from 145 to 302 ms, each tone centered within its interval. These interval series correspond to approximately linear rate ramps (fast tempo: from ≈ 4.5 Hz increasing at ≈ 3 Hz/s; slow tempo: from ≈ 7 Hz decreasing at ≈ 2 Hz/s). In abrupt trials, because tone onsets were centered within their nominal temporal periods, the first tone of the new rate occurred approximately half an interonset interval after the 1000-ms visual change point (≈ 56 ms later for the 9-Hz fast tempo rate and ≈ 167 ms later for the 3-Hz slow tempo rate).

The following schematic (Figure 2) illustrates the three visual-change modes and the gradual auditory sequences: Gradual visual profiles are linear ramps connecting the midpoints of the prechange and postchange segments of the corresponding abrupt profiles (linear change over 500–1500 ms), and the gradual fast/slow tempo sequences contain 15 tones (intervals 207–97 ms) and 10 tones (intervals 145–302 ms), respectively, spanning the 2000-ms trial.

2.1.4 | Procedure for Experiment 1

Each trial began with a 500-ms fixation period followed by the audiovisual stimulus stream. Participants performed a two-alternative forced-choice task and reported whether the visual grating became faster or slower while ignoring the sounds. Response keys (“z” and “m”) were counterbalanced across participants. Responses were accepted from the onset of the second half of the stimulus until 2000 ms after stimulus offset, followed by a 2000-ms intertrial interval. Consequently, in gradual trials a response could occur before the programmed change had fully unfolded.

Each participant completed 960 trials, with 12 repetitions per minimal condition. Trials were organized into 16 blocks of 60 trials. Abrupt and gradual visual-change conditions were blocked, and no-sound trials were administered in separate blocks. Testing was completed in two sessions of approximately

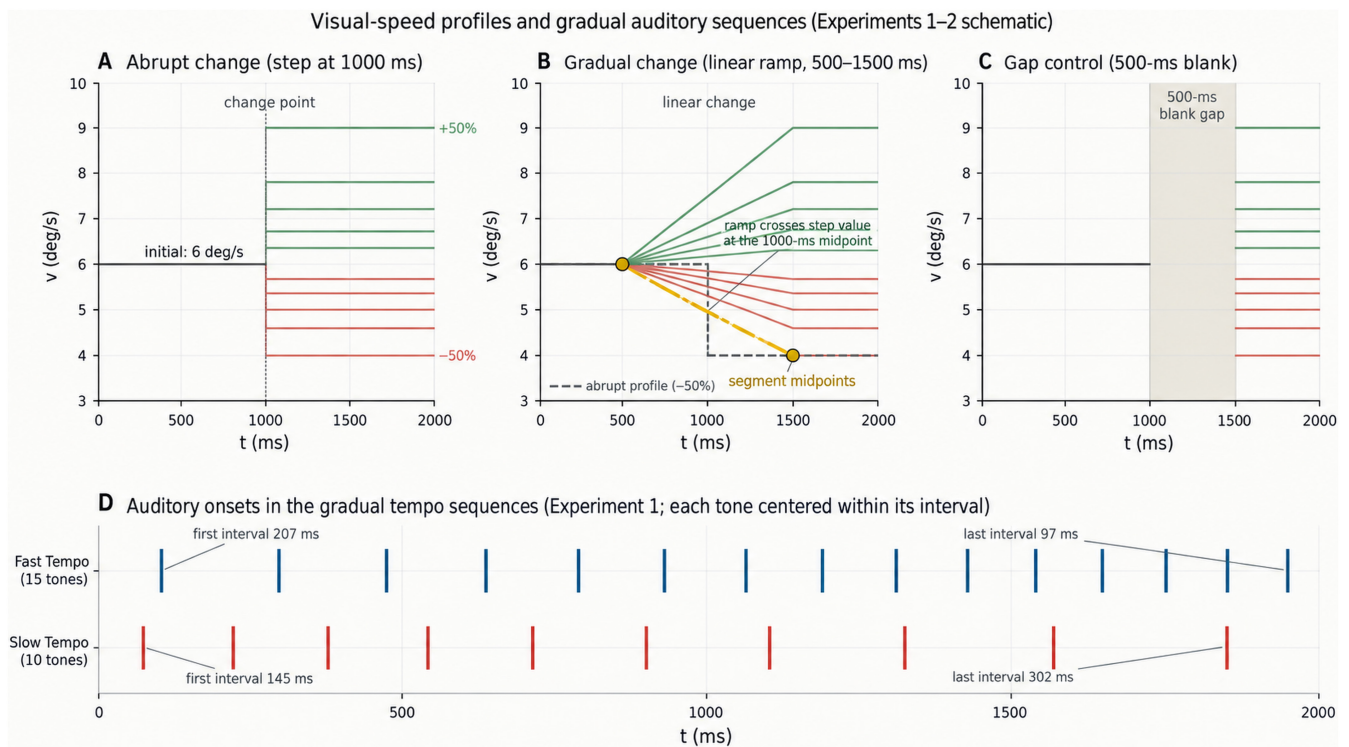


FIGURE 2 | Visual-speed profiles and gradual auditory sequences of Experiments 1 and 2. (A) Abrupt condition: speed steps from 6°/s to one of ten final speeds (4–9°/s) at the 1000-ms midpoint. (B) Gradual condition: speed remains at the initial value for 0–500 ms, changes linearly between the midpoints of the prechange and postchange segments (500–1500 ms, crossing the abrupt step value at 1000 ms; the dashed step and gold markers illustrate the midpoint-connection construction), and remains at the final value for 1500–2000 ms. (C) Gap control: a 500-ms blank separates the two motion segments. (D) Tone onsets of the Experiment 1 gradual tempo sequences: fast tempo, 15 tones with interonset intervals of 207–97 ms; slow tempo, 10 tones with intervals of 145–302 ms; each tone was centered within its interval.

50 min, separated by 4–48 h. Before the formal experiment, participants completed 24 practice trials for each sound-present change mode.

2.2 | Experiment 2: Effect of Sound Pitch on Visual Speed Perception

2.2.1 | Participants

A total of 15 students from Peking University participated in the experiment. One participant was excluded due to low accuracy. The remaining 14 valid participants (eight males and six females) had a mean age of 21.36 years ($SD = 2.31$). All participants had normal or corrected-to-normal vision and reported normal hearing. Participants read and signed an informed consent form prior to the experiment. The study was approved by the Committee for Protecting Human and Animal Subjects of the School of Psychological and Cognitive Sciences, Peking University. No participant took part in both Experiments 1 and 2. This nonoverlap was deliberate: Because the order of the two experiments could not be counterbalanced, prior participation in Experiment 1 might have influenced performance in Experiment 2, so the two experiments recruited entirely separate samples. The two samples are therefore statistically independent, and the pooled cross-experiment analysis is between-participants (see the Formal Cross-Experiment Comparison and Table S4). Sample size was based on comparable within-subject studies; a sensitivity power analysis indicated that the final sample ($N = 14$) provides 80% power to detect within-subject effects of $d_z \geq 0.81$ (two-tailed paired-samples t -test, $\alpha = 0.05$). The observed reliable participant-level PSE effects ($|dz| = 1.17$ – 1.59) exceeded this threshold.

2.2.2 | Stimuli and Apparatus

Visual stimuli and apparatus were identical to Experiment 1. Auditory sequences were isochronous at 6 Hz and consisted of twelve 50-ms sinusoidal tones with 10-ms onset and offset ramps. Pitch, rather than presentation rate, was manipulated. The reference pitch was 800 Hz. In ascending-pitch trials, pitch increased toward 1200 Hz; in descending-pitch trials, it decreased toward 500 Hz. In gradual trials, pitch changed linearly over time. Thus, Experiment 2 preserved auditory event rate while manipulating pitch trajectory (Figure 3).

2.2.3 | Procedure for Experiment 2

Experiment 2 used three visual-change modes (abrupt, gradual, or change after a blank interval) and auditory conditions defined by ascending pitch, descending pitch, and no sound.

The trial sequence and response task were identical to Experiment 1. The auditory control condition was matched to the visual-change mode: Gradual-change control trials contained no sound at all, whereas abrupt-change control trials contained a single 50-ms, 800-Hz sinusoidal tone (10-ms onset/offset ramps) presented at a random time within ± 100 ms of the visual speed

change; this brief tone provided the same timing cue that the abrupt pitch transition carried in experimental trials, thereby controlling the temporal-cue confound. Each participant completed 1080 trials (12 repetitions per minimal condition), organized into 18 blocks of 60 trials. Visual-change modes and the no-sound condition were blocked. Testing was completed in two sessions of approximately 55 min, separated by 4–48 h, with 24 sound-present practice trials per visual-change mode.

2.3 | Experiment 3: The Effect of Auditory Rhythm Changes on the Extrapolation of Uniformly Accelerated Motion

2.3.1 | Participants

A total of 25 university students participated in the experiment. Data from four participants were excluded: One because valid trials under the prespecified eye-response proximity criterion constituted less than two-thirds of the total; two because of excessive blinking during eye-movement recording; and one because of an atypical saccade pattern. The retained sample therefore comprised 21 participants (eight males and 13 females; mean age = 20.58 years, $SD = 2.32$). Three of these participants had also taken part in Experiment 2 two weeks earlier; Experiment 3 shared no experimental stimuli with Experiment 2, and participant-name records indicated no other cross-experiment overlap. All participants had normal or corrected-to-normal vision and self-reported normal hearing. All participants provided written informed consent. The study was approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences, Peking University. We checked trial-level reanalysis file used for the verified repeated-measures analyses, which contained complete condition data for 19 participants; accordingly, the inferential Experiment 3 statistics reported below use $N = 19$ ($df = 18$ for two-level paired effects). A sensitivity power analysis for $N = 19$ indicated 80% power to detect within-subject effects of $d_z \geq 0.68$ (two-tailed paired-samples t -test, $\alpha = 0.05$).

2.3.2 | Stimuli and Apparatus

The visual stimulus was a red ball (1° diameter) moving from left to right toward a black occluder (height = 2° , length = 13.4°) whose left edge was centered on the display. During the 1700-ms visible interval, the ball followed one of five constant-acceleration trajectories (-2.3 , -1.3 , 0 , 1.6 , or $3.2^\circ/s^2$) and reached the occluder at a common final visible velocity of $8^\circ/s$. Because the starting position varied with acceleration, the corresponding physical TTC values after occlusion were 2810, 2000, 1675, 1461, and 1324 ms, respectively.

Auditory stimuli were sequences of 30-ms, 500-Hz sinusoidal tones with 5-ms onset and offset ramps. The accelerating tempo sequence began at 4.5 Hz and increased at 3 Hz/s; the decelerating tempo sequence began at 7 Hz and decreased at 2 Hz/s. Both had an average rate of approximately 6 Hz during the first second. Auditory stimulation began at occlusion and ended at the participant's response, with a maximum duration of 4 s (Figure 4). Both sequences were fixed-length pregenerated

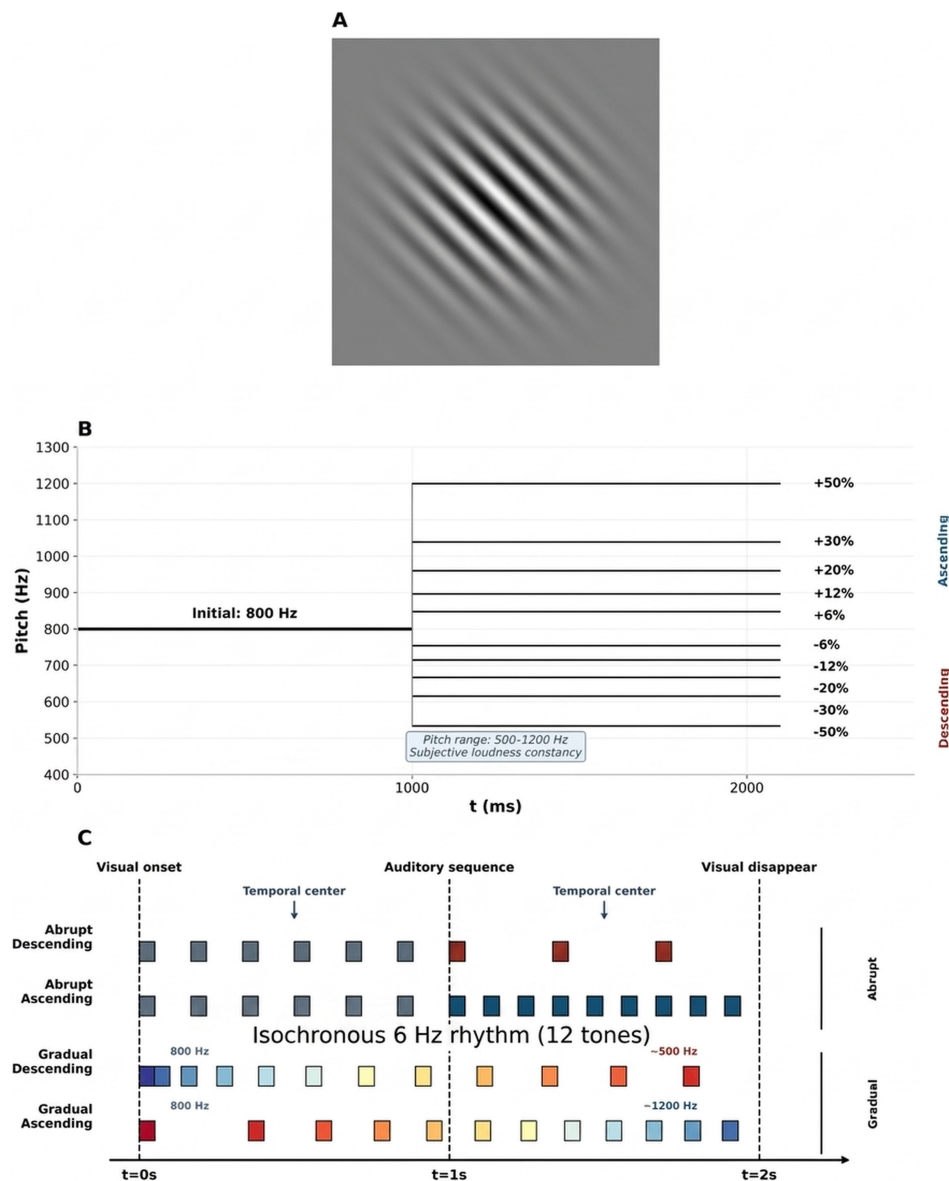


FIGURE 3 | Stimuli and setup for Experiment 2. (A) Example of the drifting Gabor stimulus. (B) Pitch profiles for the 10 signed pitch-change levels in the abrupt condition, with an 800-Hz reference and ascending or descending trajectories. (C) Temporal coupling of auditory and visual stimuli. Auditory event rate remained constant at 6 Hz while pitch either changed abruptly at the midpoint or changed continuously during gradual trials. The control condition was matched to the visual-change mode: The control was completely silent (no sound) for gradual visual-speed changes, whereas abrupt-change control trials contained a single 50-ms, 800-Hz sinusoidal tone presented at a random time within ± 100 ms of the visual change point (see Procedure for Experiment 2).

buffers rather than tones scheduled from a running rate law. The decelerating tempo buffer contained 13 tones spanning 4000 ms, with interonset intervals increasing monotonically from 145 ms to a final interval of 1000 ms (1 Hz); the nominal 0-Hz point of the linear $7-2t$ description at 3.5 s was therefore never realized, and no tones occurred beyond the 4000-ms buffer end. The accelerating tempo buffer contained 42 tones spanning approximately 3999 ms, with intervals decreasing from 207 to 61 ms. Playback began at occlusion onset and stopped immediately at the manual response, so the number of delivered tones depended on response duration: Across typical response times (approximately 1.3–2.8 s), decelerating tempo delivered 7–12 tones and accelerating tempo delivered 9–25 tones (see Table S5, for the full tone-count mapping).

2.3.3 | Procedure for Experiment 3

Experiment 3 used a 5×3 within-subjects design crossing five preocclusion physical accelerations (-2.3 , -1.3 , 0 , 1.6 , and $3.2^\circ/\text{s}^2$) with auditory condition (no sound, decelerating tempo, accelerating tempo). Earlier shorthand labels (-40% , -20% , 0% , $+20\%$, and $+40\%$) are omitted here because they do not transparently define a physical percentage quantity. The primary dependent variable was behavioral TTC; eye tracking provided complementary indices of extrapolative oculomotor dynamics.

Each trial began with a 400–800-ms fixation interval at the target's starting position. Target onset was gaze contingent (gaze

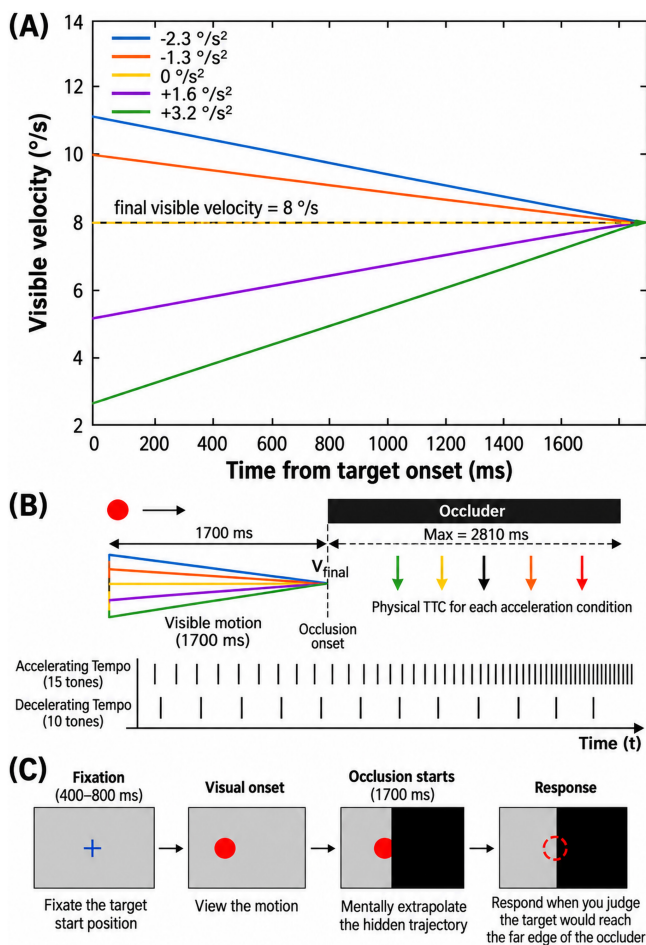


FIGURE 4 | Experiment 3 design. (A) Preocclusion visible-velocity trajectories for the five acceleration conditions (-2.3 , -1.3 , 0 , $+1.6$, and $+3.2^\circ/\text{s}^2$); the y axis shows visible velocity in $^\circ/\text{s}$, and all trajectories converge on the common final visible velocity of $8^\circ/\text{s}$ at occlusion onset (dashed line). (B) Trial timing: The target was visible for 1700 ms and disappeared behind the occluder, and auditory tempo began changing at occlusion onset. Colored arrows indicate the physical TTC associated with each preocclusion acceleration trajectory; tick marks show the accelerating and decelerating tempo tone sequences. (C) Participants fixated the target start position, viewed the motion, mentally extrapolated the hidden trajectory, and responded when they judged the target would reach the far edge of the occluder.

within 2° of fixation). After the target disappeared behind the occluder, participants mentally extrapolated its motion and pressed the spacebar when they judged that the target would reach the far edge of the occluder. The intertrial interval was 2000 ms (Figure 4). Participants received the following instructions: In each trial, they would see a ball moving from left to right with uniformly changing speed; after the ball reached the center of the screen it would be occluded but would continue to move with the same uniform acceleration behind the occluder; after the ball disappeared, they were to keep tracking it with their eyes and to press the spacebar when it reached the right edge of the occluder; the ball would not reappear, and the system would automatically record the accuracy of their tracking and of the key press; sounds would be presented during the trial, were irrelevant to task performance, but should not be completely ignored.

The experiment consisted of 360 trials in total, divided into 9 blocks of 40 trials each. Each minimal condition contained 24 trials. The silent control condition was conducted in separate blocks. The total duration was approximately 80 min. No feedback regarding estimation error was provided during the formal experiment to minimize learning effects. A practice session of 30 trials was provided beforehand, during which the ball reappeared at its true position after the button press to familiarize participants with the task.

Eye movements were recorded with a head-fixed EyeLink 1000 Plus (SR Research). Most participants were recorded binocularly at 1000 Hz; the first three participants were recorded monocularly at 1000 Hz and two participants were recorded binocularly at 500 Hz. Acquisition settings used velocity and acceleration thresholds of $22^\circ/\text{s}$ and $4000^\circ/\text{s}^2$, respectively.

3 | Analysis Methods

3.1 | EXP1 Data Preprocessing and Statistical Analysis

Primary inference used the valid Abrupt and Gradual trials. Signed visual speed change was modeled continuously, and binary acceleration responses were analyzed with binomial psychometric models on participant $\times$ signed-speed $\times$ visual-pattern $\times$ auditory-condition trial counts. All models included visual pattern, signed speed, their interaction, and participant-specific baseline and speed-slope deviations. Five nested auditory-effect models tested no auditory effect, condition-specific bias/PSE shifts, condition-specific slope/JND changes, both effects, or additional moderation by visual-change pattern. Model selection used AIC, trial-count BIC, likelihood-ratio tests, and leave-one-participant-out predictive log likelihood. PSE and JND_{75-50} were derived from the fitted intercept and slope; 95% confidence intervals were obtained by participant bootstrap, with participant-level two-parameter logistic fits retained as confirmatory summaries.

3.2 | EXP2 Data Preprocessing and Statistical Analysis

The primary analysis paralleled Experiment 1 and focused on the complete Abrupt/Gradual $\times$ no sound/descending/ascending subset. Binary acceleration responses were reconstructed from accuracy and signed speed change, and signed visual speed change was modeled continuously. The trial sequence and response task were otherwise identical to Experiment 1.

One participant was excluded because of low task performance, leaving $N=14$. No-response trials (accuracy = -1) were removed before the primary analysis. Response reconstruction and the five-model binomial psychometric analysis followed Experiment 1. Candidate models tested whether pitch trajectory shifted psychometric bias/PSE, psychometric slope/JND, both, or whether these effects varied by abrupt versus gradual visual change. AIC, trial-count BIC, likelihood-ratio tests, leave-one-participant-out prediction, participant bootstrap confidence intervals, and participant-level confirmatory fits were used in parallel with Experiment 1.

3.3 | Cross-Experiment and Exploratory Model Analyses

For the formal Experiment 1-versus-2 comparison, valid Abrupt/Gradual trials were pooled after experiment-specific preprocessing. Auditory conditions were harmonized as no sound, a slowing-direction cue (slow tempo/descending pitch), and a speeding-direction cue (fast tempo/ascending pitch). A binomial psychometric model included signed speed, visual pattern, experiment, auditory cue, and relevant interactions, with participant-specific baseline and speed-slope terms; Wald inference used participant-cluster-robust covariance. The two experiments share no participants, so the pooled analysis is between-participants. Exploratory raw-curve comparisons additionally contrasted bias-only, flexible psychometric, forced-fusion, segregation, and full Bayesian causal-inference model families on a common response likelihood; these group-aggregated fits are treated as descriptive model comparisons rather than participant-level estimates of causal-inference parameters.

3.4 | EXP3 Data Preprocessing and Statistical Analysis

3.4.1 | Eye-Movement Preprocessing

Raw gaze position, pupil samples, EyeLink events, and trial messages were parsed from each ASC file. Trial timing was defined by recorded visual-onset, occlusion/auditory-onset, and response messages, preserving the native 1000-Hz sampling rate. Gaze position was low-pass filtered at 15 Hz and pupil area at 10 Hz; short bounded gaps were interpolated, blink-contaminated periods were padded and excluded, and saccade periods were excluded from pursuit traces before derived eye velocity was low-pass filtered at 30 Hz. Saccades were classified by horizontal displacement as forward or backward. Primary eye-TTC was the elapsed time from occlusion onset to the end of the first postocclusion forward saccade landing at or beyond the collision-zone coordinate, because occluded targets

are tracked largely through predictive sequences of saccades (Tammi et al. 2022) that, together with pursuit, reflect a common sensorimotor process (Orban de Xivry and Lefèvre 2007). All inferential eye-TTC results reported in the revised manuscript use this primary definition.

3.4.2 | Dependent Variables and Statistical Analysis

We reported the analyses that directly address the central hypotheses: manual TTC, primary eye-TTC, occlusion-aligned eye velocity, pupil dynamics, and predictive TTC-model comparison. Eye-velocity and pupil traces were baseline corrected using the 200 ms immediately preceding occlusion. Time-resolved decelerating-minus-silent and accelerating-minus-silent contrasts were evaluated with participant-level sign-flip cluster-permutation tests collapsed across acceleration. Secondary forward/backward-saccade counts, intersaccadic-interval models, window-based mixed models, and eye-behavior association models are reported, together with their complete methods, in the Supporting Information.

Manual TTC and primary eye-TTC condition means were fit separately with candidate models contrasting final-velocity, average-velocity, kinematic, effective-acceleration, and tempo-dependent formulations (Table 1). The primary comparison was participant-level leave-one-condition-out (LOCO) prediction. Among the nine models, M8 was specified as $TTC = p1 \cdot TTC_{phys} + p2 \cdot (Tempo \cdot TTC_{phys}) + p3$, where $p1$ is a global physical-TTC scaling term, $p2$ is a tempo-by-physical-TTC gain, and $p3$ is an additive offset. M8 is interpreted as a phenomenological functional model; its multiplicative structure is compatible with, but does not by itself establish, an internal-clock mechanism.

The candidate set contains the following nine formulations: M1, M2a, M2b, M3a, M3b, M4a, M4b, M5, and M8. Model parameters are reported in seconds or dimensionless scaling units as appropriate to each equation.

TABLE 1 | Candidate model specifications for Experiment 3 manual and eye-based TTC fitting.

Model ID	Model name	Governing equation	Parameters
M1	Final Velocity Additive	$TTC = V_{finald} + p \cdot Tempo$	$k = 1$ (p)
M2a	Average Velocity Additive	$TTC = V_{avgd} + p \cdot Tempo$	$k = 1$ (p)
M2b	Average Velocity Multiplicative	$TTC = V_{avgd} \cdot (1 + p \cdot Tempo)$	$k = 1$ (p)
M3a	Kinematic (V_{final}) Additive	$TTC = TTC_{phys}(V_{final}, a) + p \cdot Tempo$	$k = 1$ (p)
M3b	Kinematic (V_{final}) Multiplicative	$TTC = TTC_{phys}(V_{final}, a) \cdot (1 + p \cdot Tempo)$	$k = 1$ (p)
M4a	Kinematic (V_{avg}) Additive	$TTC = TTC_{phys}(V_{avg}, a) + p \cdot Tempo$	$k = 1$ (p)
M4b	Kinematic (V_{avg}) Multiplicative	$TTC = TTC_{phys}(V_{avg}, a) \cdot (1 + p \cdot Tempo)$	$k = 1$ (p)
M5	Effective Acceleration Scaling	$TTC = TTC_{phys}(V_{avg}, a \cdot (1 + p \cdot Tempo))$	$k = 1$ (p)
M8	Three-Parameter Perception	$TTC = p1 \cdot TTC_{phys} + p2 \cdot (Tempo \cdot TTC_{phys}) + p3$	$k = 3$ ($p1, p2, p3$)

4 | Results

4.1 | Results for Experiment 1

4.1.1 | Experiment 1: Trial-Level Model Comparison and PSE

The most robust model-selection result was a bias shift without an auditory slope change. BIC and LOSO prediction selected the Bias-only model (AIC = 1896.70, BIC = 2123.14, LOSO $-\log\text{Lik}$ = 1079.26). Adding auditory bias terms to the Base model strongly improved fit, likelihood-ratio $\chi^2(2) = 228.74$, $p < 0.001$, whereas adding auditory slope terms beyond the bias model did not, $\chi^2(2) = 3.51$, $p = 0.173$. AIC was slightly lower for the Full model, but its additional pattern-specific terms worsened BIC and out-of-sample prediction. Thus, the conservative trial-level conclusion is that auditory tempo primarily shifts PSE rather than changing discrimination slope (Table 2).

Across abrupt and gradual patterns, slow tempo shifted PSE positively relative to no sound by 0.0426, 95% bootstrap CI [0.0243, 0.0639], whereas fast tempo shifted PSE negatively by -0.0587 , 95% CI $[-0.0757, -0.0449]$, yielding a slow-fast difference of 0.1013, 95% CI [0.0757, 0.1326]. The participant-level confirmatory fits showed the same pattern: slow – no sound, $t(13) = 3.84$, $p = 0.002$, $d_z = 1.03$; fast – no sound, $t(13) = -7.15$, $p < 0.001$, $d_z = -1.91$; and slow – fast, $t(13) = 6.01$, $p < 0.001$, $d_z = 1.61$. Importantly, the sign of the PSE shift indicates the direction of perceptual bias: A positive PSE means that greater physical acceleration was required to elicit 50% “acceleration” responses and thus reflects a deceleration bias, whereas a negative PSE reflects an acceleration bias. Accordingly, slow tempo induced a deceleration bias, while fast tempo induced an acceleration bias, with a similar directional pattern observed for both abrupt and gradual visual changes (Table 2 and Figure 5). Relative to the $\pm 50\%$ auditory-rate manipulation, the fitted visual PSE shifts were approximately 4%–6% of the reference speed-change scale, corresponding to roughly a 10% cross-modal transfer in magnitude. Individual-participant PSE shifts are plotted in Figure 6: In both visual-pattern conditions, the large majority of the 14 participants fell on the expected side of zero for both the slow – no sound and fast – no sound contrasts, showing that the group-level tempo effect was not driven by a minority of participants. A confirmatory GLMM with participant random effects (full coefficients in Table S3) reproduced both auditory bias effects (both $|z| > 2.7$, $p_s \leq 0.006$) and found no reliable speed $\times$ gradual interaction ($\beta = 0.043$, $SE = 0.196$, 95% CI $[-0.342, 0.427]$, $z = 0.22$, $p = 0.828$).

Pattern-specific trial-level estimates are reported in Table S1. The auditory PSE shifts were present in both abrupt and gradual conditions, and the participant bootstrap did not indicate a reversal or dependence of the main tempo effect on visual-change pattern.

4.1.2 | Experiment 1: Discrimination Precision (JND_{75-50})

In contrast to the PSE shifts, there was no reliable evidence that auditory tempo altered discrimination precision in this paradigm (Table 3). All three collapsed JND contrasts had 95% participant-bootstrap confidence intervals that included zero, and the participant-level paired comparisons were nonsignificant ($p_s \geq 0.324$). This null result is specific to the present speed-change task and stimulus regime; it should not be interpreted as evidence that auditory timing cannot affect low-level visual-motion sensitivity under other conditions (Table 4).

4.1.3 | Experiment 1 Summary

Experiment 1 therefore provides strong evidence for a bidirectional auditory-tempo effect on perceptual/decision bias, but

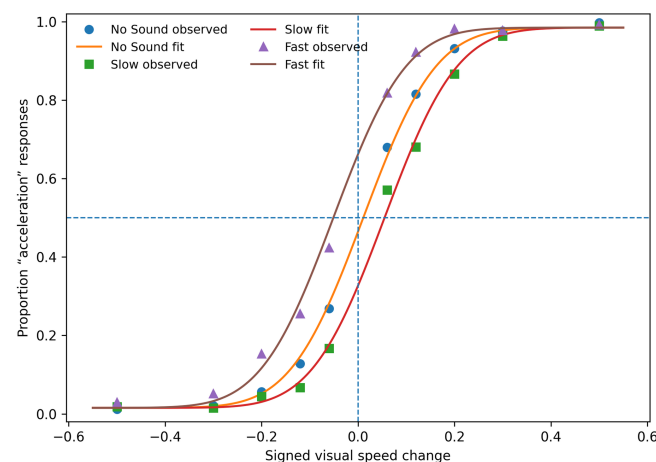


FIGURE 5 | Experiment 1 collapsed raw psychometric data and the raw-curve Bias-only model. Observed acceleration-response proportions are shown across signed visual speed change for no sound, slow tempo, and fast tempo. The principal auditory effect is a horizontal displacement of the psychometric function.

TABLE 2 | Trial-level psychometric model comparison in Experiment 1.

Model	AIC	BIC	LOSO $-\log\text{Lik}$	Interpretation
Bias-only	1896.70	2123.14	1079.26	PSE/bias shift only
Bias+Slope	1897.19	2137.79	1081.99	PSE + slope
Full	1889.97	2158.87	1082.49	Pattern moderation
Base	2121.43	2333.73	1181.89	No auditory effect
Slope-only	2113.77	2340.21	1181.22	Slope change only

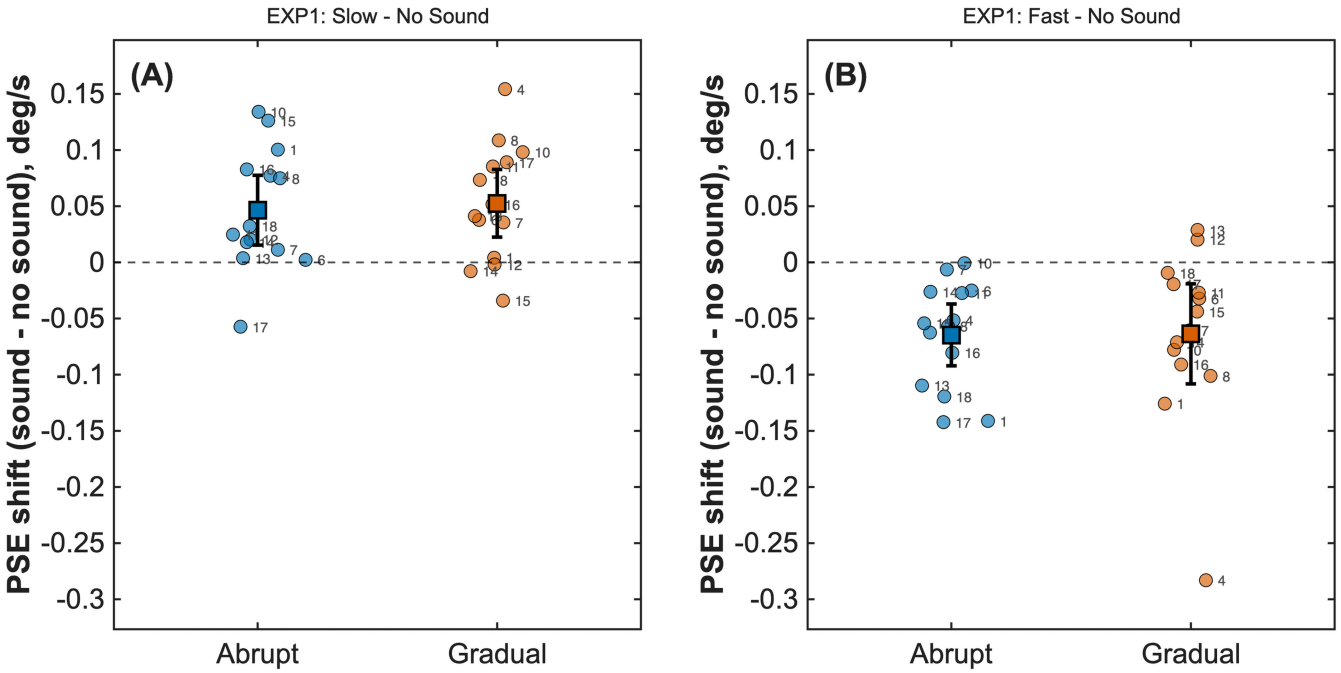


FIGURE 6 | Experiment 1 individual-participant PSE shifts (sound minus no sound). (A) Slow tempo – no sound; (B) fast tempo – no sound. Each dot is one participant (IDs labelled), shown separately for abrupt and gradual visual-change patterns; black squares show group means with 95% confidence intervals. Positive values indicate a deceleration-direction PSE shift. Per-condition PSE and JND_{75-50} estimates are reported in Table S1.

TABLE 3 | Collapsed auditory contrasts in experiment 1.

Metric	Contrast	Estimate	95% bootstrap CI	Participant-level check
PSE	Slow – no sound	0.0426	[0.0243, 0.0639]	$t(13) = 3.84, p = 0.002$
PSE	Fast – no sound	−0.0587	[−0.0757, −0.0449]	$t(13) = −7.15, p < 0.001$
PSE	Slow – fast	0.1013	[0.0757, 0.1326]	$t(13) = 6.01, p < 0.001$
JND_{75-50}	Slow – no sound	—	CI includes 0	$p \geq 0.324$
JND_{75-50}	Fast – no sound	—	CI includes 0	$p \geq 0.324$
JND_{75-50}	Slow – fast	—	CI includes 0	$p \geq 0.324$

TABLE 4 | Collapsed auditory contrasts in Experiment 2.

Metric	Contrast	Estimate	95% bootstrap CI	Participant-level check
PSE	Descending – no sound	0.0390	[0.0239, 0.0549]	$t(13) = 4.39, p < 0.001$
PSE	Ascending – no sound	−0.0094	[−0.0218, 0.0066]	$t(13) = −1.21, p = 0.250$
PSE	Descending – ascending	0.0484	[0.0329, 0.0630]	$t(13) = 5.93, p < 0.001$
JND_{75-50}	Descending – no sound	−0.0005	[−0.0103, 0.0101]	$t(13) = 0.11, p = 0.918$
JND_{75-50}	Ascending – no sound	0.0039	[−0.0039, 0.0123]	$t(13) = 0.91, p = 0.380$
JND_{75-50}	Descending – ascending	−0.0045	[−0.0154, 0.0067]	$t(13) = −0.61, p = 0.555$

little evidence for a change in visual speed-change precision. Slow tempo shifts judgements toward deceleration and fast tempo toward acceleration. Both BIC and held-out prediction favor the simpler Bias-only account, and a direct raw-curve

model comparison likewise favors a condition-specific bias model over a full Bayesian causal-inference model. The result indicates a robust tempo-dependent bias shift rather than dynamic causal uncertainty or sensitivity enhancement.

4.2 | Results for Experiment 2

4.2.1 | Experiment 2: Trial-Level Model Comparison and PSE

Pitch trajectory produced a reliable bias effect but no reliable auditory slope effect. Adding descending/ascending bias terms to the base model improved fit, $\chi^2(2)=65.99$, $p<0.001$. Adding slope terms beyond the bias-only model did not, $\chi^2(2)=1.56$, $p=0.459$. BIC and LOSO prediction selected bias-only (AIC=2075.78, BIC=2306.55, LOSO $-\log\text{Lik}=1209.70$). Although the full model had a slightly lower AIC (2071.79) and

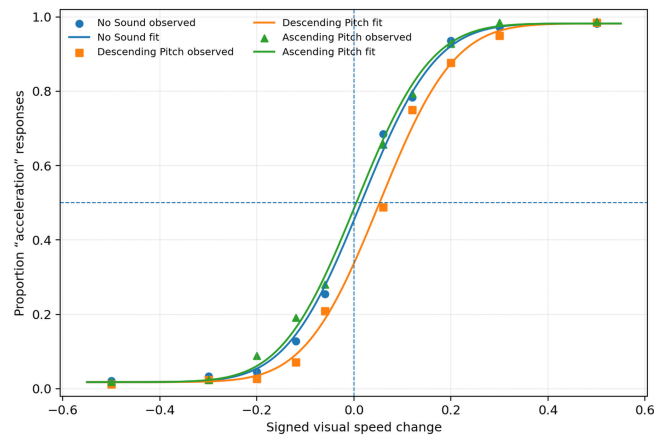


FIGURE 7 | Experiment 2 collapsed raw psychometric data and the bias-only fit. Descending pitch shifts the response curve toward a positive PSE, whereas ascending pitch remains close to no sound. Curves are collapsed across abrupt and gradual visual-change patterns for the raw-curve model comparison.

its joint pattern-moderation terms improved in-sample likelihood, it had a substantially higher BIC (2345.83) and worse held-out prediction (LOSO $-\log\text{Lik}=1211.72$). The conservative predictive conclusion is therefore a pitch-related bias shift rather than a pitch-related change in slope (Figure 7).

The visual-change pattern itself was allowed to modulate psychometric slope through the speed $\times$ gradual term. The fixed-effects trial-count fit suggested broader psychometric functions for gradual than abrupt changes, but a confirmatory GLMM with participant random effects showed that this interaction was not reliable once participant-level variability was modeled (Experiment 2: $\beta=0.090$, $\text{SE}=0.183$, 95% CI $[-0.269, 0.449]$, $z=0.49$, $p=0.624$; Experiment 1: $\beta=0.043$, $\text{SE}=0.196$, 95% CI $[-0.342, 0.427]$, $z=0.22$, $p=0.828$; full coefficients in Table S3). For these GLMMs, signed speed was recoded so that positive values denote acceleration (the opposite sign convention from the earlier analysis) and was rescaled before fitting. This pattern term was not specific to a particular pitch condition and should be distinguished from an auditory JND effect.

Descending pitch produced a positive PSE shift relative to no sound of 0.0390, 95% participant-bootstrap CI $[0.0239, 0.0549]$, corresponding to a bias toward perceived visual deceleration. Ascending pitch produced a small negative shift of -0.0094 , 95% CI $[-0.0218, 0.0066]$, which was not reliably different from zero. Descending and ascending pitch differed by 0.0484, 95% CI $[0.0329, 0.0630]$. Participant-level fits confirmed the same asymmetric pattern: descending – no sound, $t(13)=4.39$, $p<0.001$, $d_z=1.17$; ascending – no sound, $t(13)=-1.21$, $p=0.250$; descending – ascending, $t(13)=5.93$, $p<0.001$, $d_z=1.59$.

The positive descending-pitch PSE shift was present for both abrupt and gradual visual changes. Ascending pitch remained close to no

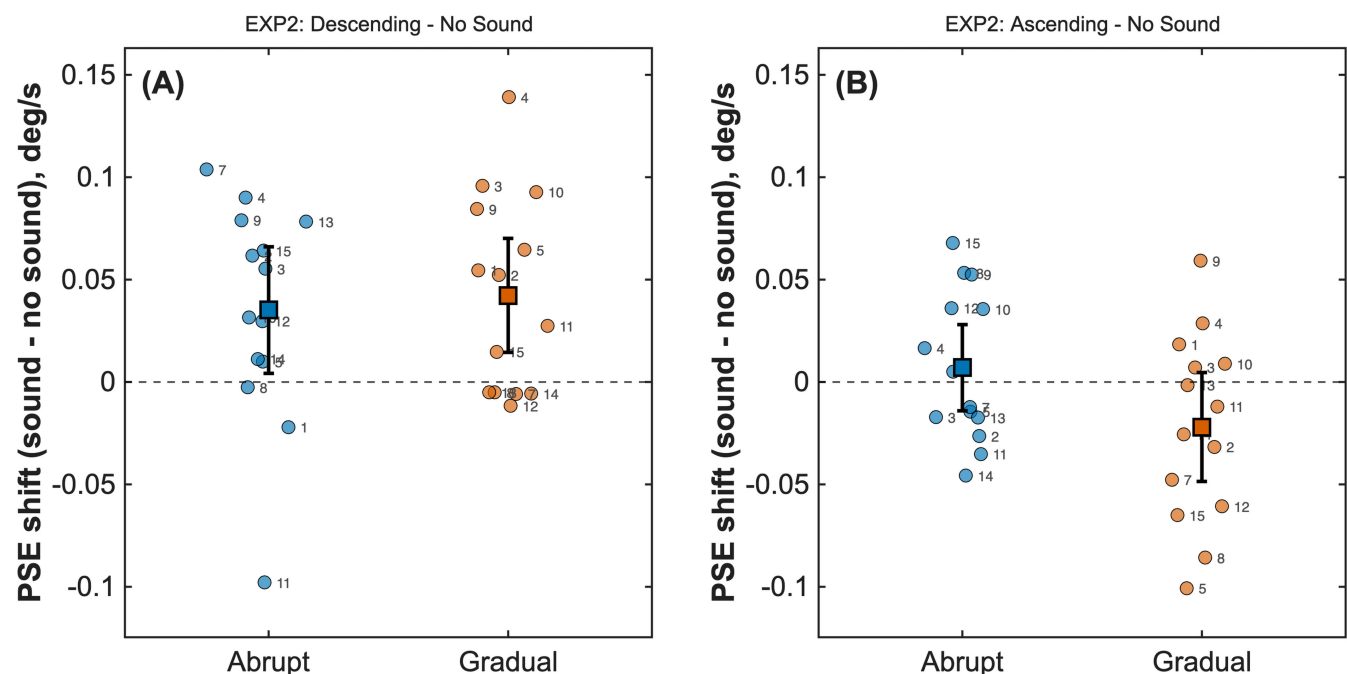


FIGURE 8 | Experiment 2 individual-participant PSE shifts (sound minus no sound). (A) Descending pitch – no sound; (B) ascending pitch – no sound. Conventions as in Figure 6. Per-condition PSE and JND75–50 estimates are reported in Table S2.

sound in both patterns. Thus, the pitch effect is asymmetric: descending pitch biases judgements toward deceleration, but ascending pitch does not provide a reliable mirror-image acceleration bias. Individual-participant shifts are plotted in Figure 8: Nearly all participants showed a positive descending – no sound shift, whereas ascending – no sound shifts were distributed around zero. The confirmatory random-effects GLMM (Table S3) agreed: The descending-cue term was reliable ($z = -2.62$, $p = 0.009$), whereas the ascending-cue term was not ($z = -1.94$, $p = 0.052$).

4.2.2 | Experiment 2: Discrimination Precision (JND₇₅₋₅₀)

The trial-level analysis does not support a pitch-related reduction in discrimination precision. Collapsed across visual pattern, descending – no sound JND = -0.0005 , 95% CI [-0.0103 , 0.0101]; ascending – no sound = 0.0039 , 95% CI [-0.0039 , 0.0123]; and descending – ascending = -0.0045 , 95% CI [-0.0154 , 0.0067]. All intervals include zero.

4.2.3 | Formal Cross-Experiment (Experiment 1 vs. Experiment 2) Trial-Level Comparison

The pooled dataset contained 28 experiment-specific participant records and 18,758 valid abrupt/gradual trials (8746 in Experiment 1 and 10,012 in Experiment 2). Auditory cues were harmonized as no sound, slowing-direction, and speeding-direction conditions as specified in the Methods. No participant took part in both experiments, so the two samples are statistically independent. The pooled GLMM therefore treats the 28 experiment-specific participant records as 28 independent participants, with participant-specific baseline and speed-slope deviations and participant-cluster-robust covariance (full coefficients in Table S4). Both harmonized auditory-cue terms remained highly significant in this identity-preserving model ($z = -3.61$ and $z = -3.75$, both $ps < 0.001$), confirming that the auditory bias effects do not depend on treating the overlapping participants as independent.

The experiment $\times$ auditory-cue interaction on psychometric bias was significant, Wald $\chi^2(2) = 25.65$, $p < 0.001$. The slowing-direction cue did not differ across experiments (Experiment 2 $\times$ slowing: $\beta = 0.085$, cluster-robust SE = 0.179 , $z = 0.48$, $p = 0.635$), whereas the speeding-direction cue was strongly attenuated in Experiment 2 (experiment 2 $\times$ speeding: $\beta = -0.673$, SE = 0.138 , $z = -4.86$, $p < 0.001$). In the BIC-preferred pooled bias-difference model, the fitted PSE shifts were $+0.0422$ (slowing) and -0.0589 (fast) in experiment 1, compared with $+0.0391$ (descending) and -0.0095 (ascending) in Experiment 2. Thus, the slowing/decelerating bias generalized across auditory dimensions, but the opposite speeding/acceleration bias was specific to the tempo manipulation.

There was no reliable cross-experiment difference in auditory modulation of psychometric slope (experiment $\times$ cue $\times$ speed: Wald $\chi^2(2) = 1.20$, $p = 0.550$), and the higher-order pattern-dependent cross-experiment terms were not jointly reliable with cluster-robust inference, Wald $\chi^2(4) = 8.95$, $p = 0.062$. BIC favored the model allowing experiment-specific auditory bias

shifts (BIC = 4512.91) over a model with shared auditory effects (4537.52), additional experiment-specific slope effects (4529.98), or the full pattern-moderated cross-experiment model (4557.55). Confirmatory participant-level contrasts yielded the same interpretation: The slowing-direction PSE shift did not differ between experiments, Welch $t(23.88) = 0.53$, $p = 0.604$, whereas the speeding-direction shift did, $t(24.89) = -4.87$, $p < 0.001$; the slow/fast versus descending/ascending directional span was also larger in Experiment 1, $t(17.76) = 3.19$, $p = 0.005$. None of the corresponding cross-experiment JND contrast differences was reliable (all $p \geq 0.418$).

4.2.4 | Exploratory Raw-Curve Computational Model Comparison (Experiment 1 vs. Experiment 2)

4.2.4.1 | Experiment 1. Bias-only provided the best information-criterion tradeoff (NLL = 84.29, AIC = 178.59, BIC = 213.97). Flexible psychometric (AIC = 181.91, BIC = 231.44), forced fusion (AIC = 181.95, BIC = 231.49), and full BCI (AIC = 183.57, BIC = 240.18) achieved only small raw-fit gains that did not offset their additional parameters. The fitted full BCI also approached a high common-cause regime and showed boundary-like decision-noise behavior, further cautioning against mechanistic interpretation.

4.2.4.2 | Experiment 2. Flexible psychometric had the lowest AIC (183.67), only 1.07 AIC units below bias-only (184.74), but bias-only had the lowest trial-likelihood BIC (220.80 vs. 234.15 for flexible psychometric). Full BCI was worse by both criteria (AIC = 186.76, BIC = 244.45), as was forced fusion (AIC = 186.07, BIC = 236.55). Thus, Experiment 2 does not provide evidence that a causal-uncertainty mechanism is needed to explain the pitch effect (Figure 9).

Taken together, the psychometric and computational model comparisons favor a simpler account in which auditory context shifts the perceptual estimate or decision criterion for visual speed change (Figure 10). The data do not rule out multisensory causal inference as a general framework, but the present Experiments 1–2 do not require a BCI mechanism and do not identify participant-level common-cause parameters.

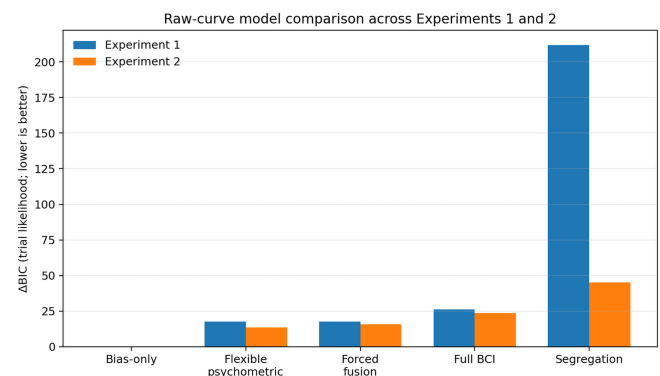


FIGURE 9 | Raw-curve model comparison across Experiments 1 and 2. Bars show Δ BIC relative to the best model within each experiment using the same binomial trial likelihood. Bias-only has the lowest BIC in both experiments; full BCI incurs a substantial complexity penalty without a compensating improvement in fit.

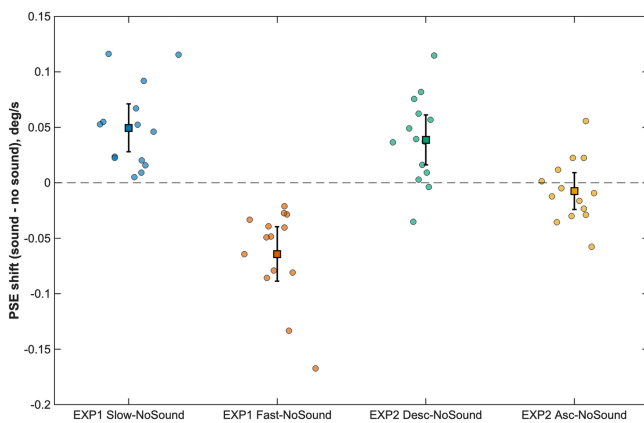


FIGURE 10 | Cross-experiment comparison of individual-participant PSE shifts. Pattern-collapsed sound – no sound PSE shifts are shown side by side for the two auditory conditions of each experiment. Each dot is one participant; black squares show group means with 95% confidence intervals. The slowing-direction shifts (slow tempo in Experiment 1, descending pitch in Experiment 2) were comparable across experiments, whereas the speeding-direction shift was markedly weaker for ascending pitch than for fast tempo, consistent with the formal trial-level contrasts reported in the text.

4.2.5 | Experiment 2 Summary

Experiment 2 showed a reliable but asymmetric pitch-trajectory effect on visual speed-change bias. Descending pitch shifted PSE toward perceived deceleration, whereas ascending pitch did not produce a reliable opposite PSE shift. Trial-level model found no reliable pitch-related JND increase; discrimination was generally poorer for gradual than abrupt visual changes, but this was a visual-pattern effect rather than an auditory effect. Thus, pitch and tempo should not be treated as interchangeable cues: Both can bias judgements, but tempo produces a stronger bidirectional mapping. The pooled comparison localized the cross-experiment difference to the speeding-direction cue: Descending pitch was statistically comparable to slow tempo, whereas ascending pitch was markedly weaker than fast tempo; no corresponding cross-experiment slope/JND difference was detected (Figure 10).

Both experiments are best characterized primarily by shifts in psychometric bias, with Experiment 1 showing two directions of tempo-induced shift and Experiment 2 showing only the descending-pitch component. Because the tempo manipulation produced the clearer bidirectional perceptual effect, Experiment 3 focused on auditory rhythm and asked whether changing tempo can also bias extrapolation after the visual target becomes occluded.

4.3 | Results for Experiment 3

4.3.1 | Experiment 3: Behavioral and Eye-TTC Estimates

In the verified repeated-measures analyses, we used the 19 participants represented in the archived trial-level reanalysis file. For manual TTC, the full 3 (auditory condition) × 5

(acceleration) analysis showed no significant three-level auditory main effect, Greenhouse–Geisser (GG) $\epsilon = 0.528$, $F(1.056, 19.009) = 0.418$, $p = 0.537$, partial $\eta^2 = 0.023$. The acceleration effect was significant with uncorrected degrees of freedom but did not remain significant after GG correction, $\epsilon = 0.319$, $F(1.276, 22.977) = 3.548$, $p = 0.063$, partial $\eta^2 = 0.165$; the auditory condition × acceleration interaction was also not significant, $\epsilon = 0.583$, $F(4.665, 83.962) = 0.277$, $p = 0.915$, partial $\eta^2 = 0.015$. In the planned comparison restricted to the two sound-present conditions, tempo was significant, $F(1, 18) = 6.383$, $p = 0.021$, partial $\eta^2 = 0.262$: Accelerating tempo shortened judged TTC by approximately 32 ms relative to decelerating tempo (approximately 1965 vs. 1997 ms). In this sound-present analysis, neither acceleration, $\epsilon = 0.332$, $F(1.327, 23.895) = 3.260$, $p = 0.073$, partial $\eta^2 = 0.153$, nor the tempo × acceleration interaction, $\epsilon = 0.791$, $F(3.164, 56.952) = 0.250$, $p = 0.870$, partial $\eta^2 = 0.014$, was significant. Thus, the manual result is a sound-present accelerating-versus-decelerating contrast, not a statistically established bidirectional shift relative to silence.

Primary eye-TTC did not show a reliable tempo effect. In the full 3 × 5 repeated-measures analysis, the auditory main effect was not significant, GG $\epsilon = 0.576$, $F(1.152, 20.732) = 0.039$, $p = 0.877$, partial $\eta^2 = 0.002$. The acceleration effect also did not remain significant after GG correction, $\epsilon = 0.352$, $F(1.409, 25.359) = 3.080$, $p = 0.079$, partial $\eta^2 = 0.146$, and the auditory condition × acceleration interaction was not significant, $\epsilon = 0.629$, $F(5.028, 90.503) = 0.634$, $p = 0.675$, partial $\eta^2 = 0.034$. Restricting the analysis to the two sound-present conditions likewise yielded no tempo effect, $F(1, 18) = 0.178$, $p = 0.678$, partial $\eta^2 = 0.010$, no GG-corrected acceleration effect, $\epsilon = 0.371$, $F(1.482, 26.682) = 2.400$, $p = 0.122$, partial $\eta^2 = 0.118$, and no tempo × acceleration interaction, $\epsilon = 0.855$, $F(3.421, 61.581) = 0.422$, $p = 0.763$, partial $\eta^2 = 0.023$. These results place the null primary eye-TTC contrast on the same inferential footing as the manual analysis.

Figure 11 plots manual TTC and primary eye-TTC against physical TTC for the three auditory conditions, with M8 predictions and the identity line. Under the primary definitions, the condition means were as follows. Silent: manual M = 1955 ms, 95% CI [1819, 2092]; primary eye-TTC M = 1208 ms, 95% CI [1088, 1328]. Decelerating tempo: manual M = 1997 ms, 95% CI [1878, 2116]; eye-TTC M = 1200 ms, 95% CI [1090, 1309]. Accelerating tempo: manual M = 1965 ms, 95% CI [1852, 2077]; eye-TTC M = 1204 ms, 95% CI [1103, 1306]. The sound-present manual contrast (accelerating – decelerating) was -32.3 ms, 95% CI $[-59.2, -5.4]$, $t(18) = -2.53$, $p = 0.021$, $d_z = -0.58$, whereas the primary eye-TTC contrast was small and not reliable (4.6 ms, 95% CI $[-18.3, 27.6]$, $t(18) = 0.42$, $p = 0.678$, $d_z = 0.10$). Both paired contrasts are based on the same $N = 19$ analysis sample used in the verified repeated-measures analyses. Primary eye-TTC estimates were substantially shorter than manual estimates, and both measures were heavily compressed relative to the identity line, indicating that extrapolation did not fully track the physical TTC manipulation (Table 5), consistent with the limit use of acceleration information in prediction-motion tasks (Benguigui and Bennett 2010; Bennett and Benguigui 2013; Vicovaro et al. 2019).

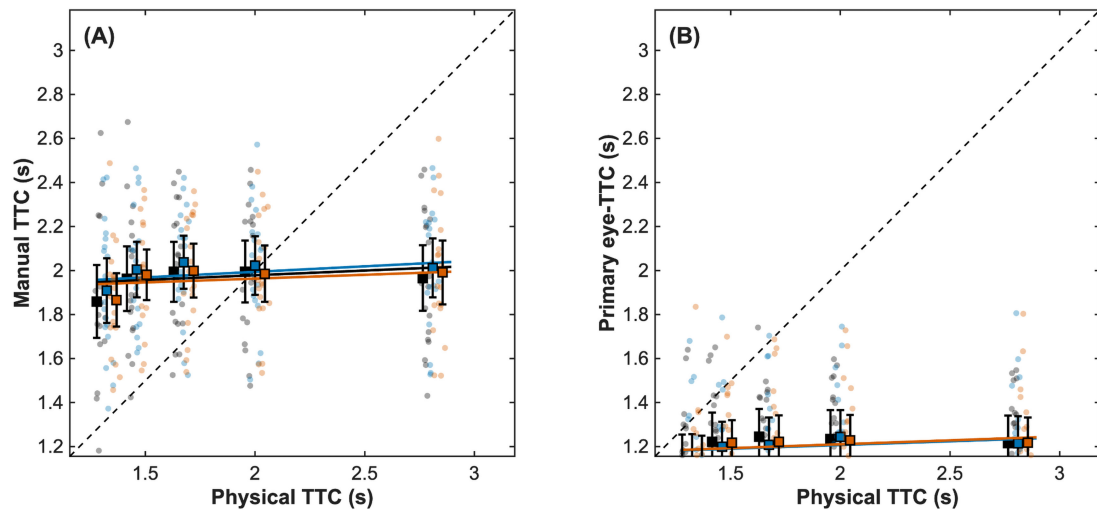


FIGURE 11 | Experiment 3 condition-level time-to-contact estimates. (A) Manual TTC estimates. (B) Primary eye-TTC estimates (first postocclusion forward saccade reaching the collision zone). Both panels plot condition means against physical TTC with 95% confidence intervals; dots show individual participants, the dashed line is the identity line, and solid curves show M8 predictions for each auditory condition.

TABLE 5 | Verified repeated-measures summary of Experiment 3 manual TTC and primary eye-TTC.

Outcome	Analysis/effect	Statistic	Key descriptive contrast
Manual TTC	3 × 5 full	Tempo: $\epsilon = 0.528$, $F(1.056, 19.009) = 0.418$, $p = 0.537$, $\eta^2 = 0.023$; acceleration: $\epsilon = 0.319$, $F(1.276, 22.977) = 3.548$, $p = 0.063$, $\eta^2 = 0.165$	Omnibus sound ns; GG acceleration $p = 0.063$; interaction $p = 0.915$
Manual TTC	2 × 5 sound-present	Tempo: $F(1, 18) = 6.383$, $p = 0.021$, $\eta^2 = 0.262$; acceleration: $\epsilon = 0.332$, $F(1.327, 23.895) = 3.260$, $p = 0.073$	Accelerating ≈ 1965 ms vs. decelerating ≈ 1997 ms (≈ -32 ms); interaction $p = 0.870$
Primary eye-TTC	3 × 5 full	Tempo: $\epsilon = 0.576$, $F(1.152, 20.732) = 0.039$, $p = 0.877$, $\eta^2 = 0.002$; acceleration: $\epsilon = 0.352$, $F(1.409, 25.359) = 3.080$, $p = 0.079$	No significant tempo or GG-corrected acceleration effect; interaction $p = 0.675$
Primary eye-TTC	2 × 5 sound-present	Tempo: $F(1, 18) = 0.178$, $p = 0.678$, $\eta^2 = 0.010$; acceleration: $\epsilon = 0.371$, $F(1.482, 26.682) = 2.400$, $p = 0.122$	Accelerating ≈ 1204 ms vs. decelerating ≈ 1200 ms (+4.6 ms); interaction $p = 0.763$

Note: All rows are based on the $N = 19$ participants represented in the archived trial-level reanalysis file. Greenhouse–Geisser corrections are reported with corrected degrees of freedom and ϵ where applicable. The F statistic is unchanged by the correction; corrected df and p values are paired from the same GG analysis. Primary eye-TTC is defined as the first postocclusion forward saccade reaching the collision zone.

4.3.2 | Experiment 3: Occlusion-Aligned Eye-Velocity Dynamics

After subtraction of the preocclusion baseline, all three tempo conditions showed a strong transient increase in eye velocity after occlusion, followed by a decline and eventual negative velocity later in the interval (Figure 12). The largest early transient occurred in the silent condition, with smaller peaks for decelerating and accelerating tempo. The participant-level cluster-permutation comparison identified a significant decelerating-minus-silent interval centered approximately 0.48–0.60 s after occlusion. No comparably extended accelerating-minus-silent cluster was evident in the plotted result. Because the analysis explicitly subtracts the preocclusion state, this postocclusion cluster provides stronger evidence for tempo-dependent predictive oculomotor dynamics.

4.3.3 | Experiment 3: Occlusion-Aligned Pupil Dynamics

Pupil responses diverged more broadly after occlusion (Figure 13). All conditions showed increasing dilation relative to the pre-occlusion baseline, but both sound conditions rose above silent beginning early in the occluded interval and remained separated for much of the next 1–1.5 s. Cluster-permutation markers indicated a broad accelerating-minus-silent cluster from shortly after occlusion to roughly 1.45 s, and a decelerating-minus-silent cluster beginning later and extending to roughly 1.5 s. The two auditory trajectories were closely similar to each other over much of this interval. This pattern supports an auditory-context effect on postocclusion physiological/task state, but the common elevation of both sound conditions cautions against interpreting pupil dilation as a direction-specific representation of tempo.

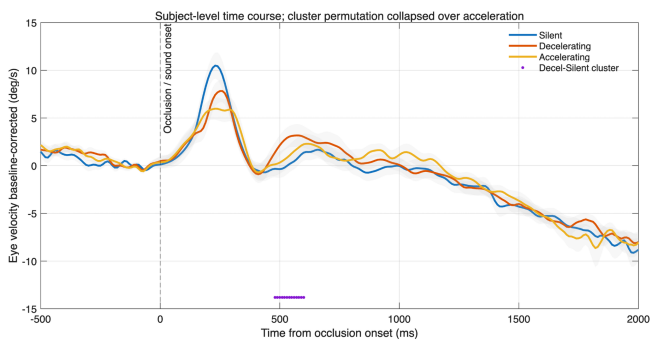


FIGURE 12 | Occlusion-aligned, baseline-corrected eye-velocity time course in Experiment 3. Curves show participant-level means collapsed across acceleration for silent, decelerating tempo, and accelerating tempo conditions; shaded bands show uncertainty across participants. Time zero marks occlusion and auditory-sequence onset. Colored markers along the lower axis denote significant cluster-permutation intervals relative to silent. The decelerating-minus-silent contrast showed a circumscribed significant cluster at approximately 480–600 ms after occlusion.

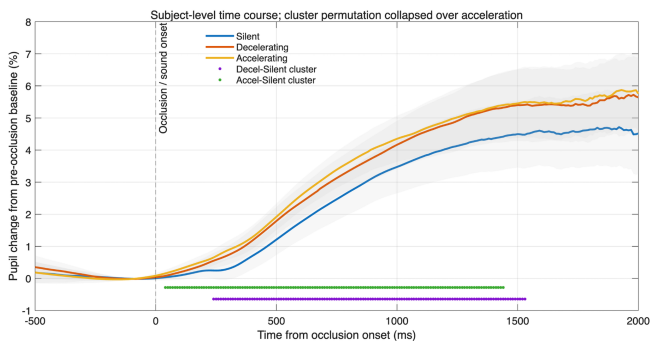


FIGURE 13 | Occlusion-aligned pupil dynamics in Experiment 3. Pupil area is expressed as percentage change from the pre-occlusion baseline and is collapsed across acceleration. Time zero marks occlusion and auditory onset. Colored markers indicate significant participant-level cluster-permutation contrasts with silent. Both auditory conditions showed sustained postocclusion dilation relative to Silent, with substantial overlap between the decelerating and accelerating trajectories.

4.3.4 | Experiment 3: Out-of-Sample TTC Model Comparison

The predictive model comparison produced a clear result (Figure 14). For manual TTC, M8 had a mean LOCO RMSE of approximately 0.17s, substantially below all other candidate models (approximately 0.39–0.65s), and was the participant-level LOCO winner for about 95% of observers. For primary eye-TTC, M8 again had the lowest mean LOCO RMSE (approximately 0.13s vs. approximately 0.52–0.98s for the alternatives) and was the LOCO winner for every participant. Thus, both response modalities are better predicted by a mapping that scales with physical TTC and allows auditory tempo to modulate that scaling than by a fixed additive tempo shift, constant-velocity heuristic, or effective-acceleration rule.

This result is predictive rather than only descriptive because each acceleration $\times$ tempo condition was held out in turn

and predicted from the remaining conditions. The tempo $\times$ physical-TTC term can be interpreted as a change in predictive weighting or temporal scaling whose influence grows with the extrapolation interval. It is compatible with a clock-rate account, but M8 is phenomenological and cannot distinguish an internal pacemaker from other mechanisms that multiplicatively transform the mapping from physical motion history to response time.

4.3.5 | Experiment 3 Summary

Experiment 3 provided a qualified set of tempo-dependent extrapolation findings. The verified full 3×5 manual analysis showed neither a significant three-level auditory effect nor a GG-corrected acceleration effect. The planned sound-present manual comparison nevertheless showed a 32-ms accelerating-versus-decelerating tempo contrast, $F(1, 18) = 6.383$, $p = 0.021$. In contrast, the corresponding primary eye-TTC tempo effect was null, $F(1, 18) = 0.178$, $p = 0.678$. Baseline-corrected eye velocity showed a temporally localized decelerating-versus-silent effect, whereas pupil dilation showed broader sound-versus-silence elevations that were similar for the two tempo directions. M8 nevertheless provided markedly better held-out prediction than the alternative TTC models for both manual and primary eye-based estimates. Thus, Experiment 3 supports a modest sound-present manual tempo contrast and tempo-related oculomotor/model signatures, but not a statistically established bidirectional TTC shift relative to silence or a corresponding primary eye-TTC tempo shift.

5 | Discussion

Across three experiments, dynamic auditory structure biased judgements of variable-speed visual motion, but the strength and form of the evidence depended on the auditory feature and task stage. Experiment 1 produced robust opposite PSE shifts for slow and fast tempo; Experiment 2 produced a weaker asymmetric pitch effect dominated by descending pitch; and Experiment 3 produced a 32-ms sound-present manual tempo contrast despite a nonsignificant omnibus three-level auditory effect. The primary eye-TTC tempo contrast in Experiment 3 was not significant, and the GG-corrected acceleration effects in both manual and primary eye-TTC analyses were also non-significant. A transient postocclusion eye-velocity effect was observed, and M8 ranked best in held-out model comparisons for both manual and eye-TTC data. These results extend prior demonstrations that sound can alter apparent motion, perceived visual speed, and inferred motion (Kafaligonul and Stoner 2010; Meyerhoff et al. 2022; Chotsrisuparat et al. 2017; Patrick and Anderson 2022; Qin et al. 2022) by showing robust auditory effects while motion is visible and more qualified tempo-dependent signatures during prediction after visual input has disappeared.

The perceptual results can be situated within causal-inference and reliability-weighted accounts without assuming that a full Bayesian causal-inference model must win the present model comparison. In Bayesian causal inference, cue combination depends on the inferred probability of a common

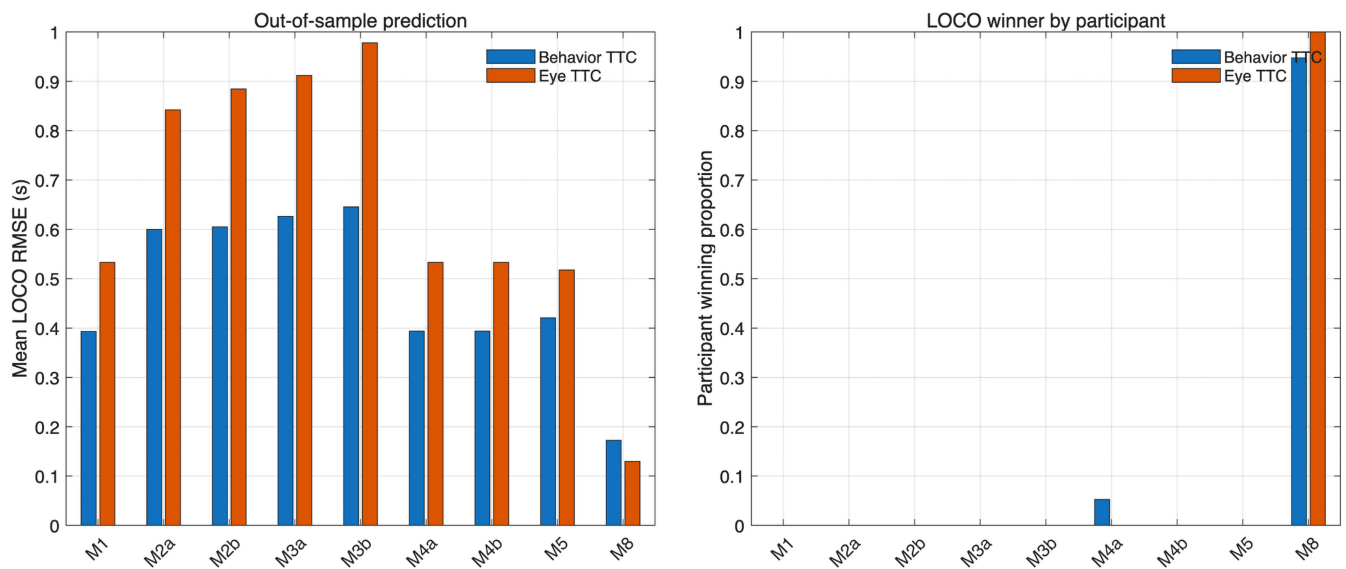


FIGURE 14 | Out-of-sample comparison of candidate TTC models for Experiment 3. Left: mean leave-one-condition-out (LOCO) RMSE for manual behavioral TTC and eye-based TTC. Right: proportion of participants for whom each model produced the lowest LOCO prediction error. M8, which combines global physical-TTC scaling, a Tempo $\times$ physical-TTC interaction, and an additive offset, clearly outperformed the alternative models for both response modalities and won for nearly all behavioral participants and all eye-TTC participants.

cause and on cue reliability (Körding et al. 2007; Shams and Beierholm 2010; Shams 2012). Visual acceleration is relatively noisy, whereas auditory timing is precise; a temporally coherent auditory sequence can therefore receive substantial effective weight when estimating a changing event. Chen et al. (2018) likewise showed graded crossmodal temporal averaging, illustrating that auditory and visual temporal signals can be partially integrated. Our raw-curve comparison found that the additional parameters of the full BCI model were not justified by the present response curves, but that negative result should be interpreted narrowly: The data do not require explicit causal-uncertainty parameters to describe these psychometric shifts, rather than disproving causal inference as a general multisensory framework.

The contrast between tempo and pitch further suggests feature-dependent weighting, even though the analysis of pitch and duration patterns may partly share neural substrates (Griffiths et al. 1999). Slow tempo and descending pitch produced comparable deceleration-directed shifts, but fast tempo produced a clear acceleration-directed shift whereas ascending pitch remained close to no sound. Global auditory rate may map more directly onto the temporal dynamics of visual speed than pitch trajectory does, and recent work showing that overall auditory interval structure, rather than only local onset/offset anchors, changes perceived biological-motion speed supports this interpretation (Özkan et al. 2026). Pitch-motion correspondences are also not necessarily symmetric: Falling pitch is ecologically associated with a receding source, and crossmodal mappings between pitch and spatial/motion dimensions can show directional asymmetries (Eitan and Granot 2006; Carnevale and Harris 2016). On the visual side, speed decrements are detected more readily than increments (Hohnsbein and Mateeff 2002), which may make deceleration-directed biases easier to obtain across auditory features. In the present stimuli, the descending and ascending excursions were also unequal in log-frequency

magnitude, so the pitch asymmetry should not be attributed to a single mechanism without a stimulus-matched follow-up.

The absence of reliable auditory JND changes in Experiments 1 and 2 should be interpreted in relation to the specific task and stimulus regime. Auditory timing can alter sensitivity to low-level visual motion in other paradigms (Kafaligonul and Stoner 2012), and task-irrelevant sound can facilitate visual processing or visual-motion representations (Kim et al. 2012; Gleiss and Kayser 2014; Kayser et al. 2017). Our results therefore support the narrower conclusion that, for these suprathreshold speed-change judgements, the dominant measurable auditory consequence was a horizontal psychometric shift rather than a slope change. Such a shift is compatible with a change in the estimated motion signal or in decision criterion; the present behavioral data do not by themselves localize the effect to a purely perceptual versus decisional stage. In audiovisual motion discrimination, for example, sensory congruency effects and history-dependent decision effects can arise at distinct processing stages (Kayser and Kayser 2018).

Experiment 3 adds to previous auditory-rhythm TTC studies by manipulating tempo only after occlusion, using multiple pre-occlusion acceleration histories that converge on the same final visible velocity, recording eye movements as an index of predictive tracking (Bennett et al. 2007; Lappi 2016; López-Moliner and Brenner 2016), and evaluating candidate models with held-out prediction. The strong performance of M8 implies that the auditory effect is not well captured by a constant additive offset. Instead, tempo interacts with physical TTC, so its influence depends on the extrapolation interval. In reliability-weighting terms, this is consistent with auditory temporal structure contributing more strongly as uncertainty accumulates during longer hidden-motion predictions. It is also compatible with the alternative that auditory tempo changes the temporal scale used to produce a TTC response rather than reweighting the visual

motion estimate itself (cf. the common rate control account of prediction motion; Makin 2018). The weaker tempo effect on primary eye- TTC than on manual TTC is also compatible with evidence that oculomotor behavior and judgments of moving or hidden targets are coupled (Jonikaitis et al. 2009; Bennett et al. 2010; Goettker et al. 2018) but can be controlled partly independently (Benguigui and Bennett 2010).

A specific version of the temporal-scale account is the internal clock or pacemaker hypothesis. Faster click trains can speed subjective timing, whereas slower temporal stimulation can produce the opposite tendency (Treisman et al. 1990; Penton-Voak et al. 1996; Wearden et al. 1999; Droit-Volet and Wearden 2002). Click trains can also increase the perceived velocity of subsequently presented moving stimuli (Makin et al. 2012). In a TTC production task, a faster clock would reach a response criterion earlier (cf. Hanes and Schall 1996) and thereby shorten judged TTC; this is the same direction observed for accelerating tempo. Scalar-timing and cortico-striatal timing frameworks provide related mechanistic context (Gibbon et al. 1984; Matell and Meck 2004; Coull et al. 2011), as do accounts linking neural oscillations to timing and temporal prediction (Gupta and Chen 2016; Wiener and Kanai 2016). The multiplicative tempo $\times$ physical-TTC structure of M8 is therefore consistent with clock-rate scaling. However, the present design does not measure subjective time independently of motion extrapolation, so it cannot adjudicate a pacemaker mechanism against other multiplicative transformations of visual prediction. A future experiment should cross auditory tempo with an independent interval-timing task or manipulate temporal reliability while holding motion evidence constant.

The eye data separate direction-specific predictive effects from more general sound-related physiology. Eye velocity showed a localized postocclusion decelerating-versus-silent effect, consistent with evidence that task-irrelevant sounds can modulate oculomotor timing (Vidal et al. 2020), whereas pupil dilation increased broadly for both sound conditions and the two auditory trajectories largely overlapped. Thus, sound per se appears to alter arousal or task state (Hess and Polt 1964; Beatty 1982), which can itself influence TTC estimation (Baures et al. 2018), while the direction of perceptual/TTC bias depends on auditory feature structure. This dissociation argues against explaining the behavioral effect as a generic arousal response. It also cautions against a single simple entrainment account (Large and Jones 1999; Lakatos et al. 2008; Henao et al. 2020; Lenc et al. 2021): A genuine entrainment mechanism would require stronger evidence of phase-specific coupling to the time-varying auditory sequence than is provided by the present oculomotor traces.

At the neural level, the findings are compatible with interactions among auditory timing systems, visual-motion cortex, and parietal predictive networks (Merchant and Georgopoulos 2006). Audiovisual synchrony can alter connectivity between auditory and visual areas (Lewis and Noppeney 2010), structural pathways connect auditory and visual motion-selective regions (Gurtubay-Antolin et al. 2021), and V5/MT+ responds to audiovisual motion events (Scheef et al. 2009) and contributes causally to spatio-temporal integration in visual motion (Zinchenko et al. 2021). Hierarchical multisensory regions can also implement computations consistent with causal inference (Rohe and Noppeney 2015).

In addition, attentional modulation in visual cortex depends on the expected timing of task events (Ghose and Maunsell 2002), offering one route through which auditory temporal structure could shape visual-motion processing. These observations identify plausible substrates, but our behavioral and eye-tracking experiments cannot determine which neural populations implement the tempo-dependent bias. Neuroimaging, electrophysiology, or causal perturbation would be required to localize the mechanism.

Several other limitations constrain the interpretation. The present pitch excursions were not perfectly symmetric in logarithmic frequency, and the TTC model comparison is phenomenological rather than a direct measurement of an internal clock. The verified Experiment 3 inferential dataset contained $N=19$ participants, for which a two-tailed paired-samples sensitivity analysis gives 80% power for effects of approximately $d_z \geq 0.68$; the observed sound-present manual contrast was $d_z = -0.58$. The supporting behavioral effect therefore falls below the effect size that this analysis sample was powered to detect and should be replicated in a larger preregistered sample. By contrast, the reliable within-subject effects in Experiments 1 and 2 (observed $|d_z| = 1.03\text{--}1.91$) exceeded their sensitivity threshold of $d_z \geq 0.81$. These issues motivate preregistered replication with matched cue magnitudes, manipulated cue reliabilities, hierarchical random-slope models, independent timing measures, and increased power for the Experiment 3 sound-present contrast.

In summary, auditory temporal structure provides a context that can bias the perception of changing visual motion and is associated with tempo-dependent signatures during extrapolation of hidden motion. The robust bidirectional tempo effect in Experiment 1, the weaker asymmetric pitch effect in Experiment 2, and the qualified sound-present/manual and model-based findings in Experiment 3 favor a feature- and context-sensitive account rather than a generic effect of sound. The data are compatible with reliability-weighted multisensory integration and with temporal-scale/clock-rate mechanisms, but they do not uniquely identify either neural or computational implementation, and Experiment 3 does not establish a bidirectional TTC shift relative to silence. Distinguishing these alternatives will require experiments that independently manipulate temporal reliability, causal correspondence, and subjective timing while measuring both perceptual and oculomotor predictions.

Author Contributions

Yantao Liu: conceptualization, data curation, investigation, formal analysis, writing – original draft, writing – review and editing. **Lihan Chen:** conceptualization, methodology, validation, formal analysis, supervision, funding acquisition, visualization, writing – original draft, writing – review and editing.

Acknowledgements

This work was supported by STI2030-Major Project (2021ZD0202600) and Natural Science Foundation of China (T2192932).

Funding

L.C. was supported by STI2030-Major Project (2021ZD0202600), and Natural Science Foundation of China (T2192932).

Ethics Statement

The study protocol was approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences, Peking University.

Conflicts of Interests

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study are publicly available at <https://osf.io/qtbjm/>.

Peer Review

For transparency, the peer review documents associated with this article are available at <https://doi.org/10.1111/ejn.70730>.

References

- Battaglini, L., G. Campana, and C. Casco. 2013. "Illusory Speed Is Retained in Memory During Invisible Motion." *Perception* 4, no. 3: 180–191. <https://doi.org/10.1068/i0562>.
- Battaglini, L., and A. Ghiani. 2021. "Motion Behind Occluder: Amodal Perception and Visual Motion Extrapolation." *Visual Cognition* 29, no. 8: 475–499. <https://doi.org/10.1080/13506285.2021.1943094>.
- Baures, R., F. Maquestiaux, P. R. DeLucia, A. Defer, and E. Prigent. 2018. "Availability of Attention Affects Time-To-Contact Estimation." *Experimental Brain Research* 236, no. 7: 1971–1984. <https://doi.org/10.1007/s00221-018-5273-8>.
- Beatty, J. 1982. "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin* 91, no. 2: 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>.
- Benguigui, N., and S. J. Bennett. 2010. "Ocular Pursuit and the Estimation of Time-To-Contact With Accelerating Objects in Prediction Motion Are Controlled Independently Based on First-Order Estimates." *Experimental Brain Research* 202, no. 2: 327–339. <https://doi.org/10.1007/s00221-009-2139-0>.
- Bennett, S. J., R. Baures, H. Hecht, and N. Benguigui. 2010. "Eye Movements Influence Estimation of Time-To-Contact in Prediction Motion." *Experimental Brain Research* 206, no. 4: 399–407. <https://doi.org/10.1007/s00221-010-2416-y>.
- Bennett, S. J., and N. Benguigui. 2013. "Is Acceleration Used for Ocular Pursuit and Spatial Estimation During Prediction Motion?" *PLoS ONE* 8, no. 5: e63382. <https://doi.org/10.1371/journal.pone.0063382>.
- Bennett, S. J., J. J. Orban de Xivry, G. R. Barnes, and P. Lefèvre. 2007. "Target Acceleration Can Be Extracted and Represented Within the Predictive Drive to Ocular Pursuit." *Journal of Neurophysiology* 98, no. 3: 1405–1414. <https://doi.org/10.1152/jn.00132.2007>.
- Brainard, D. H. 1997. "The Psychophysics Toolbox." *Spatial Vision* 10, no. 4: 433–436. <https://doi.org/10.1163/156856897X00357>.
- Brenner, E., B. Driesen, and J. B. Smeets. 2014. "Precise Timing When Hitting Falling Balls." *Frontiers in Human Neuroscience* 8: 342. <https://doi.org/10.3389/fnhum.2014.00342>.
- Brenner, E., I. A. Rodriguez, V. E. Muñoz, et al. 2016. "How Can People Be so Good at Intercepting Accelerating Objects if They Are so Poor at Visually Judging Acceleration?" *I-Perception* 7, no. 1. <https://doi.org/10.1177/2041669515624317>.
- Carnevale, M. J., and L. R. Harris. 2016. "Which Direction Is Up for a High Pitch? Multisensory Research." 29, no. 1–3: 113–132. <https://doi.org/10.1163/22134808-00002516>.
- Chen, L., X. Zhou, H. J. Müller, and Z. Shi. 2018. "What You See Depends on What You Hear: Temporal Averaging and Crossmodal Integration." *Journal of Experimental Psychology: General* 147, no. 12: 1851–1864. <https://doi.org/10.1037/xge0000487>.
- Chotsrisuparat, C., A. Koning, R. Jacobs, and R. van Lier. 2017. "Auditory Rhythms Influence Judged Time to Contact of an Occluded Moving Object." *Multisensory Research* 30, no. 7–8: 717–738. <https://doi.org/10.1163/22134808-00002592>.
- Cope, T. E., M. Grube, and T. D. Griffiths. 2012. "Temporal Predictions Based on a Gradual Change in Tempo." *Journal of the Acoustical Society of America* 131, no. 5: 4013–4022. <https://doi.org/10.1121/1.3699266>.
- Coull, J. T., R. K. Cheng, and W. H. Meck. 2011. "Neuroanatomical and Neurochemical Substrates of Timing." *Neuropsychopharmacology* 36, no. 1: 3–25. <https://doi.org/10.1038/npp.2010.113>.
- DeLucia, P. R., D. Preddy, and D. Oberfeld. 2016. "Audiovisual Integration of Time-To-Contact Information for Approaching Objects." *Multisensory Research* 29, no. 4–5: 365–395. <https://doi.org/10.1163/22134808-00002520>.
- Droit-Volet, S., and J. Wearden. 2002. "Speeding Up an Internal Clock in Children? Effects of Visual Flicker on Subjective Duration." *Quarterly Journal of Experimental Psychology. Section B, Comparative and Physiological Psychology* 55, no. 3: 193–211. <https://doi.org/10.1080/02724990143000252>.
- Eitan, Z., and R. Y. Granot. 2006. "How Music Moves: Musical Parameters and Listeners' Images of Motion." *Music Perception* 23, no. 3: 221–248. <https://doi.org/10.1525/mp.2006.23.3.221>.
- Ernst, M. O., and M. S. Banks. 2002. "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion." *Nature* 415, no. 6870: 429–433. <https://doi.org/10.1038/415429a>.
- Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39, no. 2: 175–191. <https://doi.org/10.3758/BF03193146>.
- Ghose, G. M., and J. H. R. Maunsell. 2002. "Attentional Modulation in Visual Cortex Depends on Task Timing." *Nature* 419, no. 6907: 616–620. <https://doi.org/10.1038/nature01057>.
- Gibbon, J., R. M. Church, and W. H. Meck. 1984. "Scalar Timing in Memory." *Annals of the New York Academy of Sciences* 423: 52–77. <https://doi.org/10.1111/j.1749-6632.1984.tb23417.x>.
- Gleiss, S., and C. Kayser. 2014. "Oscillatory Mechanisms Underlying the Enhancement of Visual Motion Perception by Multisensory Congruency." *Neuropsychologia* 53: 84–93. <https://doi.org/10.1016/j.neuropsychologia.2013.11.005>.
- Goettker, A., D. I. Braun, A. C. Schütz, and K. R. Gegenfurtner. 2018. "Execution of Saccadic Eye Movements Affects Speed Perception." *Proceedings of the National Academy of Sciences* 115, no. 9: 2240–2245. <https://doi.org/10.1073/pnas.1704799115>.
- Graber, E., and T. Fujioka. 2019. "Endogenous Expectations for Sequence Continuation After Auditory Beat Accelerations and Decelerations Revealed by P3a and Induced beta-Band Responses." *Neuroscience* 413: 11–21. <https://doi.org/10.1016/j.neuroscience.2019.06.010>.
- Griffiths, T. D., I. Johnsrude, J. L. Dean, and G. G. Green. 1999. "A Common Neural Substrate for the Analysis of Pitch and Duration Pattern in Segmented Sound?" *Neuroreport* 10, no. 18: 3825–3830. <https://doi.org/10.1097/00001756-199912160-00019>.
- Gupta, D. S., and L. H. Chen. 2016. "Brain Oscillations in Perception, Timing and Action." *Current Opinion in Behavioral Sciences* 8: 161–166. <https://doi.org/10.1016/j.cobeha.2016.02.021>.
- Gurtubay-Antolin, A., C. Battal, C. Maffei, et al. 2021. "Direct Structural Connections Between Auditory and Visual Motion-Selective Regions in

- Humans." *Journal of Neuroscience* 41, no. 11: 2393–2405. <https://doi.org/10.1523/JNEUROSCI.1552-20.2021>.
- Hanes, D. P., and J. D. Schall. 1996. "Neural Control of Voluntary Movement Initiation." *Science* 274, no. 5286: 427–430. <https://doi.org/10.1126/science.274.5286.427>.
- Henao, D., M. Navarrete, M. Valderrama, and M. Le Van Quyen. 2020. "Entrainment and Synchronization of Brain Oscillations to Auditory Stimulations." *Neuroscience Research* 156: 271–278. <https://doi.org/10.1016/j.neures.2020.03.004>.
- Hess, E. H., and J. M. Polt. 1964. "Pupil Size in Relation to Mental Activity During Simple Problem-Solving." *Science* 143, no. 3611: 1190–1192. <https://doi.org/10.1126/science.143.3611.1190>.
- Hidaka, S., W. Teramoto, Y. Sugita, Y. Manaka, S. Sakamoto, and Y. Suzuki. 2011. "Auditory Motion Information Drives Visual Motion Perception." *PLoS ONE* 6, no. 3: e17499. <https://doi.org/10.1371/journal.pone.0017499>.
- Hohnsbein, J., and S. Mateeff. 2002. "Speed Decrements Are Seen Better Than Increments Through Small Apertures." *Experimental Brain Research* 142, no. 3: 434–438. <https://doi.org/10.1007/s00221-001-0964-x>.
- Hubbard, T. L. 2015. "The Varieties of Momentum-Like Experience." *Psychological Bulletin* 141, no. 6: 1081–1119. <https://doi.org/10.1037/bul0000016>.
- Jonikaitis, D., H. Deubel, and C. de' Sperati. 2009. "Time Gaps in Mental Imagery Introduced by Competing Saccadic Tasks." *Vision Research* 49, no. 17: 2164–2175. <https://doi.org/10.1016/j.visres.2009.05.021>.
- Kafaligonul, H., and G. R. Stoner. 2010. "Auditory Modulation of Visual Apparent Motion With Short Spatial and Temporal Intervals." *Journal of Vision* 10, no. 12: 31. <https://doi.org/10.1167/10.12.31>.
- Kafaligonul, H., and G. R. Stoner. 2012. "Static Sound Timing Alters Sensitivity to Low-Level Visual Motion." *Journal of Vision* 12, no. 11: 2. <https://doi.org/10.1167/12.11.2>.
- Kaya, U., and H. Kafaligonul. 2019. "Cortical Processes Underlying the Effects of Static Sound Timing on Perceived Visual Speed." *NeuroImage* 199: 194–205. <https://doi.org/10.1016/j.neuroimage.2019.05.062>.
- Kayser, S. J., and C. Kayser. 2018. "Trial by Trial Dependencies in Multisensory Perception and Their Correlates in Dynamic Brain Activity." *Scientific Reports* 8: 3742. <https://doi.org/10.1038/s41598-018-22137-8>.
- Kayser, S. J., M. G. Philiastrides, and C. Kayser. 2017. "Sounds Facilitate Visual Motion Discrimination via the Enhancement of Late Occipital Visual Representations." *NeuroImage* 148: 31–41. <https://doi.org/10.1016/j.neuroimage.2017.01.010>.
- Kim, R., M. A. K. Peters, and L. Shams. 2012. "0 + 1 > 1: How Adding Noninformative Sound Improves Performance on a Visual Task." *Psychological Science* 23, no. 1: 6–12. <https://doi.org/10.1177/0956797611420662>.
- King, C., and S. L. Prime. 2019. "Auditory Pitch Glides Influence Time-To-Contact Judgements of Visual Stimuli." *Experimental Brain Research* 237, no. 8: 1907–1917. <https://doi.org/10.1007/s00221-019-05561-8>.
- Kleiner, M., D. H. Brainard, and D. G. Pelli. 2007. "What's New in Psychtoolbox-3?" *Perception* 36, no. ECVF Abstract Supplement: 1–16. <https://doi.org/10.1177/03010066070360S101>.
- Körding, K. P., U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams. 2007. "Causal Inference in Multisensory Perception." *PLoS ONE* 2, no. 9: e943. <https://doi.org/10.1371/journal.pone.0000943>.
- Kreyenmeier, P., L. Kämmer, J. Fooker, and M. Spering. 2022. "Humans Can Track but Fail to Predict Accelerating Objects." *eNeuro* 9, no. 5: ENEURO.0185-22.2022. <https://doi.org/10.1523/ENEURO.0185-22.2022>.
- Lakatos, P., G. Karmos, A. D. Mehta, I. Ulbert, and C. E. Schroeder. 2008. "Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection." *Science* 320, no. 5872: 110–113. <https://doi.org/10.1126/science.1154735>.
- Lakens, D. 2013. "Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs." *Frontiers in Psychology* 4: 863. <https://doi.org/10.3389/fpsyg.2013.00863>.
- Lappi, O. 2016. "Eye Movements in the Wild: Oculomotor Control, Gaze Behavior & Frames of Reference." *Neuroscience & Biobehavioral Reviews* 69: 49–68. <https://doi.org/10.1016/j.neubiorev.2016.06.006>.
- Large, E. W., and M. R. Jones. 1999. "The Dynamics of Attending: How People Track Time-Varying Events." *Psychological Review* 106, no. 1: 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>.
- Lenc, T., H. Merchant, P. E. Keller, H. Honing, M. Varlet, and S. Nozaradan. 2021. "Mapping Between Sound, Brain and Behaviour: Four-Level Framework for Understanding Rhythm Processing in Humans and Non-Human Primates." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 376, no. 1835: 20200325. <https://doi.org/10.1098/rstb.2020.0325>.
- Lewis, R., and U. Noppeney. 2010. "Audiovisual Synchrony Improves Motion Discrimination via Enhanced Connectivity Between Early Visual and Auditory Areas." *Journal of Neuroscience* 30, no. 37: 12329–12339. <https://doi.org/10.1523/JNEUROSCI.5745-09.2010>.
- López-Moliner, J., and E. Brenner. 2016. "Flexible Timing of Eye Movements When Catching a Ball." *Journal of Vision* 16, no. 5: 13. <https://doi.org/10.1167/16.5.13>.
- López-Moliner, J., and S. Soto-Faraco. 2007. "Vision Affects How Fast We Hear Sounds Move." *Journal of Vision* 7, no. 12: 6–6.7. <https://doi.org/10.1167/7.12.6>.
- Maeda, F., R. Kanai, and S. Shimojo. 2004. "Changing Pitch Induced Visual Motion Illusion." *Current Biology* 14, no. 23: R990–R991. <https://doi.org/10.1016/j.cub.2004.11.018>.
- Makin, A. D. J. 2018. "The Common Rate Control Account of Prediction Motion." *Psychonomic Bulletin & Review* 25, no. 5: 1784–1797. <https://doi.org/10.3758/s13423-017-1403-8>.
- Makin, A. D., E. Poliakoff, J. Dillon, A. Perrin, T. Mullet, and L. A. Jones. 2012. "The Interaction Between Duration, Velocity and Repetitive Auditory Stimulation." *Acta Psychologica* 139, no. 3: 524–531. <https://doi.org/10.1016/j.actpsy.2012.01.013>.
- Matell, M. S., and W. H. Meck. 2004. "Cortico-Striatal Circuits and Interval Timing: Coincidence Detection of Oscillatory Processes." *Cognitive Brain Research* 21, no. 2: 139–170. <https://doi.org/10.1016/j.cogbrainres.2004.06.012>.
- Merchant, H., and A. P. Georgopoulos. 2006. "Neurophysiology of Perceptual and Motor Aspects of Interception." *Journal of Neurophysiology* 95, no. 1: 1–13. <https://doi.org/10.1152/jn.00422.2005>.
- Meyerhoff, H. S., N. A. Gehrer, S. Merz, and C. Frings. 2022. "The Beep-Speed Illusion: Non-Spatial Tones Increase Perceived Speed of Visual Objects in a Forced-Choice Paradigm." *Cognition* 219: 104978. <https://doi.org/10.1016/j.cognition.2021.104978>.
- Orban de Xivry, J. J., and P. Lefèvre. 2007. "Saccades and Pursuit: Two Outcomes of a Single Sensorimotor Process." *Journal of Physiology* 584, no. 1: 11–23. <https://doi.org/10.1113/jphysiol.2007.139881>.
- Özkan, S., S. K. Yılmaz, and H. Kafalıgönül. 2026. "Biological Motion Is Perceived as Faster With Shorter Auditory Time Intervals." *Scientific Reports* 16: 28372. <https://doi.org/10.1038/s41598-026-58369-2>.
- Patrick, T. B., and R. B. Anderson. 2022. "The Influence of Auditory Rhythms on the Speed of Inferred Motion." *Attention, Perception, & Psychophysics* 84, no. 7: 2360–2383. <https://doi.org/10.3758/s13414-021-02364-4>.

- Pelli, D. G. 1997. "The VideoToolbox Software for Visual Psychophysics: Transforming Numbers Into Movies." *Spatial Vision* 10, no. 4: 437–442. <https://doi.org/10.1163/156856897X00366>.
- Penton-Voak, I. S., H. Edwards, A. Percival, and J. H. Wearden. 1996. "Speeding Up an Internal Clock in Humans? Effects of Click Trains on Subjective Duration." *Journal of Experimental Psychology: Animal Behavior Processes* 22, no. 3: 307–320. <https://doi.org/10.1037/0097-7403.22.3.307>.
- Pouliot, M., and S. Grondin. 2005. "A Response-Time Approach for Estimating Sensitivity to Auditory Tempo Changes." *Music Perception* 22, no. 3: 389–399. <https://doi.org/10.1525/mp.2005.22.3.389>.
- Qin, K. Y., W. X. Chen, J. Y. Cui, et al. 2022. "The Influence of Time Structure on Prediction Motion in Visual and Auditory Modalities." *Attention, Perception, & Psychophysics* 84, no. 6: 1994–2021. <https://doi.org/10.3758/s13414-021-02369-z>.
- Rankin, S. K., E. W. Large, and P. W. Fink. 2009. "Fractal Tempo Fluctuation and Pulse Prediction." *Music Perception* 26, no. 5: 401–413. <https://doi.org/10.1525/mp.2009.26.5.401>.
- Recanzone, G. H. 2003. "Auditory Influences on Visual Temporal Rate Perception." *Journal of Neurophysiology* 89, no. 2: 1078–1093. <https://doi.org/10.1152/jn.00706.2002>.
- Rohe, T., and U. Noppeney. 2015. "Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception." *PLoS Biology* 13, no. 2: e1002073. <https://doi.org/10.1371/journal.pbio.1002073>.
- Sadaghiani, S., J. X. Maier, and U. Noppeney. 2009. "Natural, Metaphoric, and Linguistic Auditory Direction Signals Have Distinct Influences on Visual Motion Processing." *Journal of Neuroscience* 29, no. 20: 6490–6499. <https://doi.org/10.1523/JNEUROSCI.5437-08.2009>.
- Scheef, L., H. Boecker, M. Daamen, et al. 2009. "Multimodal Motion Processing in Area V5/MT: Evidence From an Artificial Class of Audio-Visual Events." *Brain Research* 1252: 94–104. <https://doi.org/10.1016/j.brainres.2008.10.067>.
- Shams, L. 2012. "Early Integration and Bayesian Causal Inference in Multisensory Perception." In *The Neural Bases of Multisensory Processes*, edited by M. M. Murray and M. T. Wallace, 217–231. CRC Press. <https://doi.org/10.1201/9781439812174-16>.
- Shams, L., and U. R. Beierholm. 2010. "Causal Inference in Perception." *Trends in Cognitive Sciences* 14, no. 9: 425–432. <https://doi.org/10.1016/j.tics.2010.07.001>.
- Tammi, T., J. Pekkanen, S. Tuhkanen, L. Oksama, and O. Lappi. 2022. "Tracking an Occluded Visual Target With Sequences of Saccades." *Journal of Vision* 22, no. 1: 9. <https://doi.org/10.1167/jov.22.1.9>.
- Teramoto, W., S. Hidaka, J. Gyoba, and Y. Suzuki. 2010. "Auditory temporal cues can modulate visual representational momentum." *Attention, Perception, & Psychophysics* 72, no. 8: 2215–2226. <https://doi.org/10.3758/BF03196696>.
- Treisman, M., A. Faulkner, P. L. N. Naish, and D. Brogan. 1990. "The Internal Clock: Evidence for a Temporal Oscillator Underlying Time Perception With Some Estimates of Its Characteristic Frequency." *Perception* 19, no. 6: 705–743. <https://doi.org/10.1068/p190705>.
- van der Steen, M. C., N. Jacoby, M. T. Fairhurst, and P. E. Keller. 2015. "Sensorimotor Synchronization With Tempo-Changing Auditory Sequences: Modeling Temporal Adaptation and Anticipation." *Brain Research* 1626: 66–87. <https://doi.org/10.1016/j.brainres.2015.01.053>.
- Vicovaro, M., S. Noventa, and L. Battaglini. 2019. "Intuitive Physics of Gravitational Motion as Shown by Perceptual Judgment and Prediction-Motion Tasks." *Acta Psychologica* 194: 51–62. <https://doi.org/10.1016/j.actpsy.2019.02.001>.
- Vidal, M., A. Desantis, and L. Madelain. 2020. "Irrelevant Auditory and Tactile Signals, but Not Visual Signals, Interact With the Target Onset and Modulate Saccade Latencies." *PLoS ONE* 15, no. 2: e0221192. <https://doi.org/10.1371/journal.pone.0221192>.
- Watamaniuk, S. N., and S. J. Heinen. 2003. "Perceptual and Oculomotor Evidence of Limitations on Processing Accelerating Motion." *Journal of Vision* 5, no. 11: 698–709. <https://doi.org/10.1167/3.11.5>.
- Wearden, J. H., K. Philpott, and T. Win. 1999. "Speeding up and (Relatively) Slowing Down an Internal Clock in Humans." *Behavioural Processes* 46, no. 1: 63–73. [https://doi.org/10.1016/S0376-6357\(99\)00004-2](https://doi.org/10.1016/S0376-6357(99)00004-2).
- Welch, R. B., L. D. DuttonHurt, and D. H. Warren. 1986. "Contributions of Audition and Vision to Temporal Rate Perception." *Perception & Psychophysics* 39, no. 4: 294–300. <https://doi.org/10.3758/BF03204939>.
- Wessels, M., S. Kröling, and D. Oberfeld. 2022. "Audiovisual Time-To-Collision Estimation for Accelerating Vehicles: The Acoustic Signature of Electric Vehicles Impairs Pedestrians' Judgments." *Transportation Research Part F: Traffic Psychology and Behaviour* 91: 191–212. <https://doi.org/10.1016/j.trf.2022.09.023>.
- Wessels, M., C. Zähme, and D. Oberfeld. 2023. "Auditory Information Improves Time-To-Collision Estimation for Accelerating Vehicles." *Current Psychology* 42, no. 27: 23195–23205. <https://doi.org/10.1007/s12144-022-03375-6>.
- Wiener, M., and R. Kanai. 2016. "Frequency Tuning for Temporal Perception and Prediction." *Current Opinion in Behavioral Sciences* 8: 1–6. <https://doi.org/10.1016/j.cobeha.2016.01.001>.
- Zinchenko, A., S. Brunner, L. Chen, Z. Shi, P. C. J. Taylor, and H. J. Müller. 2021. "V5/MT+ Modulates Spatio-Temporal Integration Differently Across and Within Hemifields: Causal Evidence From TMS." *Neuropsychologia* 161: 107995. <https://doi.org/10.1016/j.neuropsychologia.2021.107995>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Experiment 1: auditory-condition PSE and JND₇₅₋₅₀ estimates. Collapsed contrasts across abrupt and gradual visual-change patterns (trial-level binomial model with participant bootstrap; participant-level two-parameter logistic fits as confirmatory checks). Positive PSE shifts indicate a deceleration-direction bias. **Table S2:** Experiment 2: auditory-condition PSE and JND₇₅₋₅₀ estimates. Conventions as in Table S1. **Table S3:** Participant random-effects GLMM sensitivity analyses (Experiments 1 and 2). **Table S4:** Formal pooled cross-experiment comparison (Experiment 1 vs. Experiment 2). **Table S5:** Experiment 3: pregenerated auditory-sequence buffers and delivered tone counts.