Scientific Reports

https://doi.org/10.1038/s41598-026-58369-2

Article in Press

Biological motion is perceived as faster with shorter auditory time intervals

Received: 30 January 2026

Accepted: 12 June 2026

Published online: 21 June 2026

Cite this article as: Ozkan S., Yilmaz S.K. & Kafaligonul H. Biological motion is perceived as faster with shorter auditory time intervals. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-58369-2

Sudenur Ozkan, Seyma Koc Yilmaz & Hulusi Kafaligonul

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Biological Motion Is Perceived as Faster with Shorter Auditory Time Intervals

Sudenur Ozkan[1,*], Seyma Koc Yilmaz[2], Hulusi Kafaligonul[1, 3]

[1]Department of Neuroscience, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Türkiye

[2]Institute for Cognitive Neurology and Dementia Research, Medical Faculty, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

[3]Neuroscience and Neurotechnology Center of Excellence (NÖROM), Department of Anatomy, Faculty of Medicine, Gazi University, Ankara, Türkiye

ARTICLE IN PRESS

Corresponding Author:

Sudenur Ozkan
Aysel Sabuncu Brain Research Center
Bilkent University, Ankara 06800, Türkiye
Tel.: +90 312 290 1154
Fax: +90 312 290 3001
Email: sudenur.ozkan@bilkent.edu.tr

Abstract

Perceptual experience is continuously refined by crossmodal interactions in time. Prior research shows that auditory timing can modulate the perceived speed of visual motion with simple stimuli (e.g., flashes, clicks). It remains unknown whether such temporal interactions extend to complex, socially meaningful motion, where timing carries biological significance. Here, we investigated whether auditory timing affects the perceived speed of human locomotion depicted by point-light walkers (PLWs). Across experiments,

ARTICLE IN PRESS

participants compared the walking speed of sequentially presented, visually identical PLWs accompanied by step sounds differing only in inter-sound interval. Perceived walking speed systematically rescaled with auditory timing, revealing a robust crossmodal temporal influence. To assess whether this effect relied on specific temporal anchors, we manipulated either the onset or offset of footstep timing. The crossmodal influence was not locked to onset or offset cues, indicating that overall auditory time intervals shape visual motion perception rather than acting at discrete moments of audiovisual alignment. Moreover, the effect persisted across semantically congruent and incongruent sounds, suggesting that auditory timing primarily governs perceptual alignment across modalities. Together, these findings indicate that audiovisual interactions in time extend to naturalistic biological motion. Auditory timing provides a dominant low-level cue supporting perceptual coherence in dynamic multisensory environments.

Keywords: biological motion, speed perception, multisensory, audiovisual interactions, temporal processing

ARTICLE IN PRESS

1. Introduction

Motion and speed estimation are fundamental for interacting with dynamic environments. In natural settings, moving objects rarely appear in isolation or within a single sensory modality. Instead, motion signals are typically accompanied by temporally structured information from multiple modalities. A growing number of studies have therefore emphasized the inherently multisensory nature of motion perception[1,2]. Given that moving objects typically generate characteristic sounds, auditory input can substantially influence visual motion perception, indicating that motion-related computations are not restricted to the visual domain alone[3,4]. Extensive research has demonstrated that the timing of brief static sounds (e.g., clicks) can modulate different aspects of apparent motion perception. Manipulations of auditory time intervals have been shown to bias perceived motion

ARTICLE IN PRESS

direction[5], speed[6,7], and motion categorization[8-10]. These effects are commonly interpreted in terms of auditory dominance in temporal processing and are often framed within the concept of temporal ventriloquism, in which temporally proximal sounds alter the perceived timing of visual events[3,11,12]. In the context of apparent motion, auditory clicks are thought to shift the perceived timing of successive visual motion frames or the time interval between them, thereby reshaping the final motion percept. These findings underscore the role of auditory timing in shaping motion perception and motivate the study of audiovisual interactions in dynamic contexts.

Importantly, the influence of auditory timing on visual motion perception does not appear to depend critically on attentional tracking or motion mechanisms based on tracking the positions of visual elements over time[9]. Prior work has shown that auditory temporal modulation persists when attentional demands are minimized and remains robust across variations in attentional load and the number of moving objects[13] (see also Yilmaz & Kafaligonul[14]). Converging neural evidence further supports the involvement of distributed cortical processes. EEG studies have revealed audiovisual interactions in both early and later components over parieto-occipital, parietal, and prefrontal regions[15] (see also Akdogan et al.[16]; Kaya et al.[17]), suggesting that auditory timing influences motion perception through multistage neural dynamics rather than a single locus of integration. Despite these advances, a notable limitation of existing literature is its reliance on highly simplified stimuli, typically consisting of brief visual flashes and static sounds such as clicks. While such paradigms have been instrumental in uncovering fundamental principles of audiovisual processing in the temporal domain, they are limited in addressing how these mechanisms operate under more naturalistic conditions. Bridging this gap requires extending current findings to stimuli with higher ecological validity, particularly those involving structured, meaningful motion.

ARTICLE IN PRESS

Biological motion (BM), characterized by the movement patterns of living beings, offers a compelling test case for addressing this challenge. BM carries high ecological and evolutionary relevance and is efficiently extracted even from sparse visual inputs such as point-light displays[18] (for a review see Blake & Shiffrar[19]). A growing body of research has demonstrated that the detection of BM can be systematically shaped by concurrent auditory cues, particularly when those cues are temporally, spatially, or semantically aligned with the observed action. For example, visual sensitivity to point-light displays of human walkers is enhanced when the accompanying sounds resemble natural footsteps rather than artificial tones[20,21], and when the auditory motion direction matches that of the walker[22,23]. Saygin et al.[24] further showed that the temporal binding of audiovisual signals differs between upright and inverted BM, despite their matched low-level visual properties, highlighting the role of meaningful kinematic features in crossmodal perception. Collectively, these findings suggest that BM perception engages mechanisms that are sensitive to the temporal structure and biological relevance of audiovisual signals, potentially offering crossmodal facilitation beyond simple sensory enhancement. On the other hand, existing BM studies have predominantly focused on the detection and discrimination of BM, leaving unresolved whether auditory timing can modulate perceived motion dynamics, such as speed. In particular, it remains unclear whether audiovisual interactions in BM are subject to temporal ventriloquism effects (well-documented with simple forms of stimulation) when the motion is ecologically valid.

The present study was designed to address this gap by examining whether auditory temporal structure influences perceived speed in biological motion. Previous work indicates that audiovisual temporal processing is not invariant across stimulus conditions, but can be shaped by the structural and dynamic properties of biological motion, without necessarily implying a unitary, exclusive integration mechanism. For instance, Saygin et al.[24] demonstrated

ARTICLE IN PRESS

that audiovisual temporal relations depend on the configurational integrity of biological motion, as temporal binding differed between upright point-light walkers and their inverted counterparts, even though low-level motion statistics were preserved. This finding suggests that characteristic biological kinematics modulate how temporal information is integrated across modalities. In a complementary vein, Arrighi et al.[25] reported that the temporal window for audiovisual synchrony is broader for biological motion than for artificial constant-speed motion, indicating that temporal tolerance in multisensory integration can vary as a function of motion structure and ecological relevance. Together, these findings converge on the idea that audiovisual temporal integration is sensitive to the dynamic organization of biological motion, rather than being governed solely by low-level temporal coincidence or stimulus energy. Importantly, however, prior studies have focused primarily on temporal alignment, synchrony, or detection-related measures, leaving open whether such sensitivity extends to the perception of continuous motion attributes, such as speed. Addressing this question requires moving beyond judgments of simultaneity or detectability to examine how auditory timing may reshape the perceived dynamics of biological motion itself. To this end, we tested whether manipulating auditory temporal intervals alters perceived speed in point-light walkers while holding visual input constant. In our paradigm, pairs of consecutive point-light walkers were presented with identical visual timing, and only the temporal interval between accompanying step sounds was varied. This design isolates the influence of auditory temporal structure on speed perception and allows us to assess whether temporal ventriloquism effects, which are well documented for simplified apparent motion stimuli, generalize to ecologically valid biological motion.

In the second step, we examined whether the magnitude of these auditory timing effects depends on higher-level properties of the auditory signal. Previous work has shown that semantic congruence between auditory and

ARTICLE IN PRESS

visual information can selectively modulate sensitivity to biological motion, but this selective effect depends on the level of visual awareness[26], suggesting a role for higher-level interpretative factors. By contrast, theoretical and empirical accounts of temporal ventriloquism emphasize partial-cue integration driven primarily by temporal proximity, with featural or semantic correspondence playing a secondary role[27,28]. To adjudicate between these possibilities, we extended the speed discrimination paradigm by varying the semantic content and contextual relevance of the auditory stimuli, rather than relying exclusively on footstep sounds. This approach allows us to test whether auditory temporal effects on perceived biological motion speed are constrained by semantic congruence or instead reflect more general temporal integration principles.

2. Experiment 1a

Using motion stimuli that primarily engage early stages of motion processing, previous research has demonstrated that the effects of auditory temporal intervals on visual motion perception are not confined to higher stages of sensory processing, suggesting the inherent nature of audiovisual interactions in the temporal domain[9]. Complementary EEG work has further shown that auditory timing modulates neural activity across distinct cortical regions, including occipital, parietal, and frontal areas[15-17], indicating a distributed temporal coupling mechanism within the perceptual hierarchy. Despite these insights, it remains unclear whether such temporal interactions extend to motion signals that recruit higher-level visual representations. As an initial step in the current study, we investigated whether such temporal interactions generalize to complex and ecologically valid stimuli, specifically biological motion accompanied by step sounds. To this end, we adopted a well-established speed discrimination paradigm[6,7] and paired point-light walkers (PLWs) with naturalistic footstep sounds to test whether auditory timing recalibrates the perceived speed of biological motion.

ARTICLE IN PRESS

2.1. Methods

2.1.1. Participants

Thirty-two undergraduate students (15 females; age range: 18-24 years) took part in the first three experiments (Exp. 1a, 1b, 2a). All participants had normal or corrected-to-normal vision and normal hearing, and no history of neurological disorders by self-report. Prior to participation, volunteers were informed about the experimental procedures, potential risks, compensation, and confidentiality, and signed the informed consent. Participants received course credit for their participation. An a priori power analysis (G*Power 3.1.9.7)[29] for a repeated-measures ANOVA (medium effect size f = 0.25, α = .05, power = .95) indicated a minimum sample size of 31. This sample size was determined based on the experimental condition requiring the largest number of participants to ensure the study's robustness. Accordingly, 32 participants were tested. In the final analysis, we opted for a Generalized Linear Mixed Model (GLMM) to accommodate the observed non-normal distribution of the data and to leverage trial-level variance, thereby ensuring robust parameter estimation within the determined sample size. All procedures were in accordance with the principles outlined in the Declaration of Helsinki[30] and received approval from the local Ethics Committee at Bilkent University (Approval ID: 2024.10.01.01).

ARTICLE IN PRESS

2.1.2. Apparatus

Stimulus generation, experimental design, and data acquisition were implemented in MATLAB R2023a (The MathWorks, Natick, MA, USA) using Psychtoolbox 3.0[31-33]. Visual stimuli were displayed on a gamma-corrected 27-inch LCD monitor with a 120 Hz refresh rate (1920 x 1080 pixel resolution). Screen luminance was measured using a SpectroCAL photometer (Cambridge Research Systems, Rochester, Kent, UK). A two-channel multimedia loudspeaker (Logitech Z150, Model S-00134) was used to deliver auditory stimuli. The speakers were positioned centrally beneath the monitor

ARTICLE IN PRESS

with minimal inter-speaker separation to approximate a single sound source. A digital sound-level meter (SL-4010, Lutron Electronics, Taipei, TW) was used to measure the sound pressure level. The duration of visual and auditory stimuli and their alignment were confirmed using a photodiode connected to a digital oscilloscope (DS1204B, Rigol Technologies EU, GmbH, Gilching, Germany). A chin rest was used to constrain head movements and maintain a stable viewing distance of 57 cm. A standard keyboard was used to acquire behavioral responses. The experiment was carried out in a silent and dark room. The same apparatus and testing environment were used for all experiments.

2.1.3. Stimuli and procedure

The design was based on comparing the walking speed of two sequential PLWs moving in the same (rightward) direction (Fig. 1). A white fixation circle (diameter = 0.16°, luminance = 83.16 cd/m^2) was presented at the center of the screen against a dark background (luminance = 0.21 cd/m^2). The point-light display of the walking human (i.e., the biological motion) was chosen from the Carnegie Mellon University's motion capture database (CMU Graphics Lab)[34]. Motion data were imported and visualized using the Motion Kinematic and Kinetic Analyser (Mokka)[35]. The walking motion depicted two steps (i.e., one gait cycle) comprising 120 frames and lasting 1 second. Each PLW consisted of 41 white dots (diameter = 0.05°, luminance 83.16 cd/m^2) and was presented in the upper visual field with the feet located at the center of the screen. Unlike many previous studies using sparser configurations with typically 11–15 dots[21,24,36-39], 41 were employed here to preserve fine-grained joint information and enhance ecological validity. Importantly, this choice does not alter the task structure but provides a richer biological motion signal more representative of natural human movement. Point-light walkers subtended approximately 7.68° (height) x 1.72° (width) of visual angle. Different walking speeds were created by selectively omitting motion frames between steps while inserting additional frames at the beginning and end of

ARTICLE IN PRESS

the sequence to maintain a constant stimulus duration of 1s. Frame omission was implemented uniformly across the sequence to preserve smooth motion perception and avoid local temporal artifacts (see also section 2.1.4).

Auditory stimuli consisted of naturalistic footstep sounds obtained from online databases and processed in Audacity (version 3.4.2)[40] using RMS normalization to standardize the loudness of each to 70 dB SPL. All sound files were resampled to 48 kHz. This normalization ensured consistent auditory perception across trials, minimizing variability due to amplitude differences. Only noise-free recordings were selected to maintain stimulus quality.



Figure 1. Schematic illustration of design and stimulus timing in Experiment 1a. Each trial consisted of a test (left) and a reference (right) point-light walker (PLW) sequence, each lasting 1000 ms and separated by a 500 ms blank interval. In the reference sequence, 75-ms lasting visual footsteps and sounds were temporally synchronized, with a visual inter-step interval (ISI_V) of 425 ms and a corresponding auditory inter-step interval (ISI_A) of 425 ms. In the test sequence, the auditory intervals (ISI_A = 365, 395, 455, or 485 ms) varied, yielding ΔISI_A values (differences relative to the reference) of −60,

ARTICLE IN PRESS

−30, +30, and +60 ms. Specifically, in the 425 ms ISI_A condition, the stimulus pairings for the test and reference were identical, serving as the baseline since there was no temporal shift (ΔISI_A = 0 ms [test - reference]). After viewing the two stimuli sequentially, participants were asked to judge which walker appeared to move faster.

In the present design, multiple auditory conditions were interleaved within the same experimental block. As a result, estimating separate psychometric functions for each auditory condition relative to a visual-only baseline would require a substantially larger number of trials and was not feasible within the current framework. Instead, we employed a comparative judgment approach based on within-trial comparisons of identical visual stimuli paired with different auditory intervals. This approach allows sensitive detection of relative changes in perceived speed while minimizing contextual and adaptation effects across conditions. Previous work has shown that such relative measures reliably reflect systematic shifts in perceived speed that are consistent with shifts observed in psychometric functions[6,7,13-15].

Each trial began with a white fixation point appearing in the middle of the screen for 500 ms, followed by two successive PLWs (1s each) separated by a 500 ms interstimulus interval (Fig. 1). After the presentation of the second motion, an empty screen was presented to indicate a response was required to continue on the next trial. Participants were asked to indicate, via a keypress, whether the first or second walker appeared to move faster. Crucially, participants received no specific guidance regarding the criteria for assessing speed differences, such as focusing exclusively on foot movements, arm swings, or other individual body segments. Instead, they were required to base their judgments on the overall motion and the configuration of the entire body. As in previous studies[6,7,13], one of the two successive PLWs was used as the reference and the other as the test stimulus. Participants were not informed about the actual speed of the PLWs, which did not differ from each other in terms of visual motion parameters. Instead, perceived speed differences were induced solely by manipulating time

ARTICLE IN PRESS

intervals between paired step sounds (Fig. 1). In each trial, both PLWs were visually identical (visual step interval = 425 ms), but the temporal interval between paired step sounds differed between the reference and test. The reference PLW always had auditory step intervals synchronized with its visual steps (425 ms), whereas the test PLW's auditory intervals varied across four main conditions (365, 395, 455, 485 ms), yielding ΔISI_A values (differences relative to the reference) of −60, −30, +30, and +60 ms. Each participant completed 108 trials (24 trials per ΔISI_A condition plus 12 baseline trials with $\Delta ISI_A = 0$).

2.1.4. Speed discrimination screening

Before the main experiment, each participant completed practice sessions to ensure reliable sensitivity to differences in walking speed. This session served as a screening procedure for inclusion in the main experiment. As in the main experiment, participants indicated by a keypress which PLW (first or second) appeared faster (Supplemental Fig. S1). One of the two successive PLWs was used as the reference, and the other as the test stimulus. While the visual time interval between two steps in the reference stimulus was held constant ($ISI_V = 425$ ms), it was varied for the test stimulus ($ISI_V = 350$, 375, 425, 475, and 500 ms). Participants completed 100 trials (5 Test ISI_V conditions x 20 repetitions). There was no auditory stimulus in this session, and all other stimulus parameters and procedures were the same as those described above.

For each participant, the percentage of trials in which the reference PLW was reported as faster was computed for each visual time interval condition. This value was expected to be below 30% for the shortest test interval and above 70% for the longest test interval, ensuring that participants' responses spanned a sufficient dynamic range for reliable psychometric fitting. These criteria are based on established procedures used in our previous work[13,14]. Psychometric curves were fitted to these data using a complementary error

ARTICLE IN PRESS

function (equivalent to a cumulative Gaussian) implemented in the *psignifit* MATLAB toolbox (version 4)[41]. Within the Bayesian framework implemented by psignifit, posterior distributions were estimated for the psychometric function parameters, from which threshold and slope estimates were obtained as maximum-a-posteriori (MAP) values together with their 95% credible intervals. These estimates were used solely to evaluate the quality of psychometric fits during screening. Participants were included in the main experiment only if their psychometric functions yielded valid fits across the full range of test intervals. Participants who failed to meet these predetermined criteria in three runs were excluded. Group-averaged psychometric functions are shown in Supplemental Figures S2 and S3.

2.1.5. Data Analysis

Trial-level binary responses ("test faster" vs. "reference faster") were analyzed using generalized linear mixed-effects models (GLMMs) with a binomial logit link. Analyses were conducted using the lme4 package[42] in R (version 4.4.3)[43] and RStudio (version 2024.12.1.563)[44]. This approach appropriately models binary outcomes at the trial level while accounting for the repeated-measures structure of the data and inter-individual variability. Model selection followed standard parsimony criteria based on likelihood-ratio tests and Akaike information criterion (AIC). The final model included the auditory interval (ΔISI_A, scaled in 30-ms steps, or 15-ms for Experiment 1b) as a continuous fixed effect and by-participant random slopes for ΔISI_A. Statistical inference was based on Wald *z*-tests, and planned pairwise contrasts were Holm-corrected for multiple comparisons using the *emmeans* package[45].

The GLMM was fit to all trials, including baseline trials in which the auditory intervals were identical for the reference and test walkers ($\Delta ISI_A = 0$). Baseline trials were included as observations at $\Delta ISI_A = 0$, allowing the model to estimate the full relationship between ΔISI_A and binary "test faster"

ARTICLE IN PRESS

responses across the entire range of the predictor. Although baseline trials were included in model fitting, baseline individual means are shown separately as diamonds in the figures for visualization purposes.

2.2. Results

The proportion values for different timing conditions of footstep sound are shown in Figure 2. A generalized linear mixed-effects model revealed a robust influence of auditory step timing on the perceived speed of biological motion. Increasing the auditory inter-step interval significantly decreased the likelihood that participants judged the test walker as moving faster ($\beta = -0.02$, $SE = 0.001$, $z = -19.14$, $p < 0.001$). This corresponds to an odds ratio of 0.62 (95% CI [0.59, 0.65]), indicating that each 30-ms increase in the auditory interval reduced the odds of a “test faster” judgment by approximately 38%. Predicted probabilities derived from the model showed a clear monotonic relationship. The participants were most likely to report the test as faster when the step-sound interval was 60 ms shorter (0.74 [0.71–0.76]) and least likely when it was 60 ms longer (0.29 [0.27–0.32]) than the reference (Figure 2). The intercept did not differ from zero ($\beta = 0.07$, $p = 0.08$), indicating no systematic response bias at synchrony ($\Delta ISI_A = 0$). All non-zero interval conditions differed significantly from the 0-ms baseline after Holm adjustment (all ps < 0.0001). These results demonstrate that auditory temporal intervals systematically bias the perceived speed of visually identical point-light walkers, providing clear evidence for strong crossmodal temporal interactions in biological motion perception.

ARTICLE IN PRESS

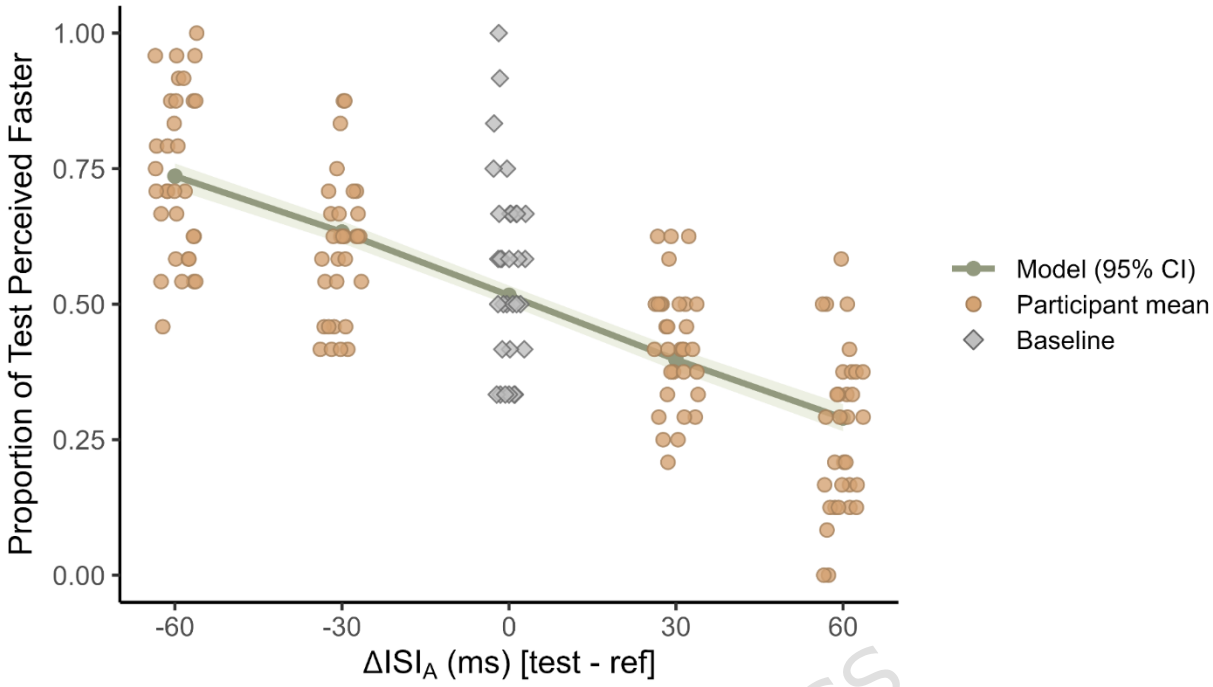


Figure 2. Proportion of "test faster" judgments as a function of auditory step-interval differences (ΔISI_A) in Experiment 1a. Each orange-filled circle represents a participant's mean proportion of "test faster" responses for each ΔISI_A condition, and gray diamonds indicate individual means for the baseline condition (identical step intervals; ΔISI_A = 0 ms). Negative ΔISI_A values correspond to shorter auditory intervals (i.e., footstep sounds for the test walker were faster), and positive values correspond to longer intervals (i.e., footstep sounds for the test walker were slower). The green line and green-filled circles represent the population-level predicted probabilities from the binomial generalized linear mixed model (GLMM), with shaded areas indicating the model's 95% confidence interval. The GLMM revealed a robust main effect of ΔISI_A: each 30-ms increase in auditory interval reduced the odds of a "test faster" judgment by approximately 38%.

3. Experiment 1b

In the classical temporal ventriloquism paradigm, the second sound (i.e., the one following the second flash) accounted for most of the temporal ventriloquism effect, particularly when its sound-light interval was up to 200 ms[12]. By contrast, the first sound exerted only a limited effect. This pattern suggests that the perceptual system may integrate crossmodal signals over a brief, asymmetric temporal window, such that later-arriving auditory

ARTICLE IN PRESS

information can retrospectively influence the perceived timing of earlier visual events. Indeed, the temporal window for audiovisual integration is shown to be asymmetric. Auditory signals slightly lagging behind a visual event tend to be integrated more readily than those that lead it[46,47]. Building on these findings, Experiment 1b examined whether a comparable asymmetry characterizes multisensory integration in biological motion perception. Specifically, we tested whether changes in perceived walking speed are driven preferentially by the timing of the first or the second footstep sound. By selectively manipulating the temporal position of the auditory shift relative to visual footsteps, we assessed how retrospective temporal binding differentially weights early versus late auditory events when integrating audiovisual cues to biological motion.

3.1. Methods

3.1.1. Participants

The same thirty-two volunteers who took part in Experiment 1a also participated in Experiment 1b, completing the experiments in separate sessions. The order of experimental sessions was counterbalanced across participants.

ARTICLE IN PRESS

3.1.2. Stimuli and procedure

The PLW stimuli, footstep sounds, and timeline of test and reference PLW during a trial were the same as those used in Experiment 1a. The participants similarly compared the speed of the test and reference PLWs, each accompanied by footstep sounds that varied in timing (Fig. 3). There were four auditory time interval conditions (ISI_A = 395, 410, 440, and 455 ms). These conditions produced two “in” and two “out” sound conditions in which the auditory intervals were either 15 ms or 30 ms shorter or longer than the corresponding visual inter-step interval (ISI_V). To dissociate the contributions of early versus late auditory information, we manipulated which sound was

ARTICLE IN PRESS

temporally shifted, yielding onset (first sound shifted) and offset (second sound shifted) conditions. This resulted in eight experimental conditions (2 [onset/offset] × 4 main [auditory intervals]) plus a synchronous baseline condition. Participants completed 12 trials per condition for a total of 108 trials. All other stimulus parameters, experimental procedures, and performance criteria (including practice and exclusion thresholds) were identical to Experiment 1a.

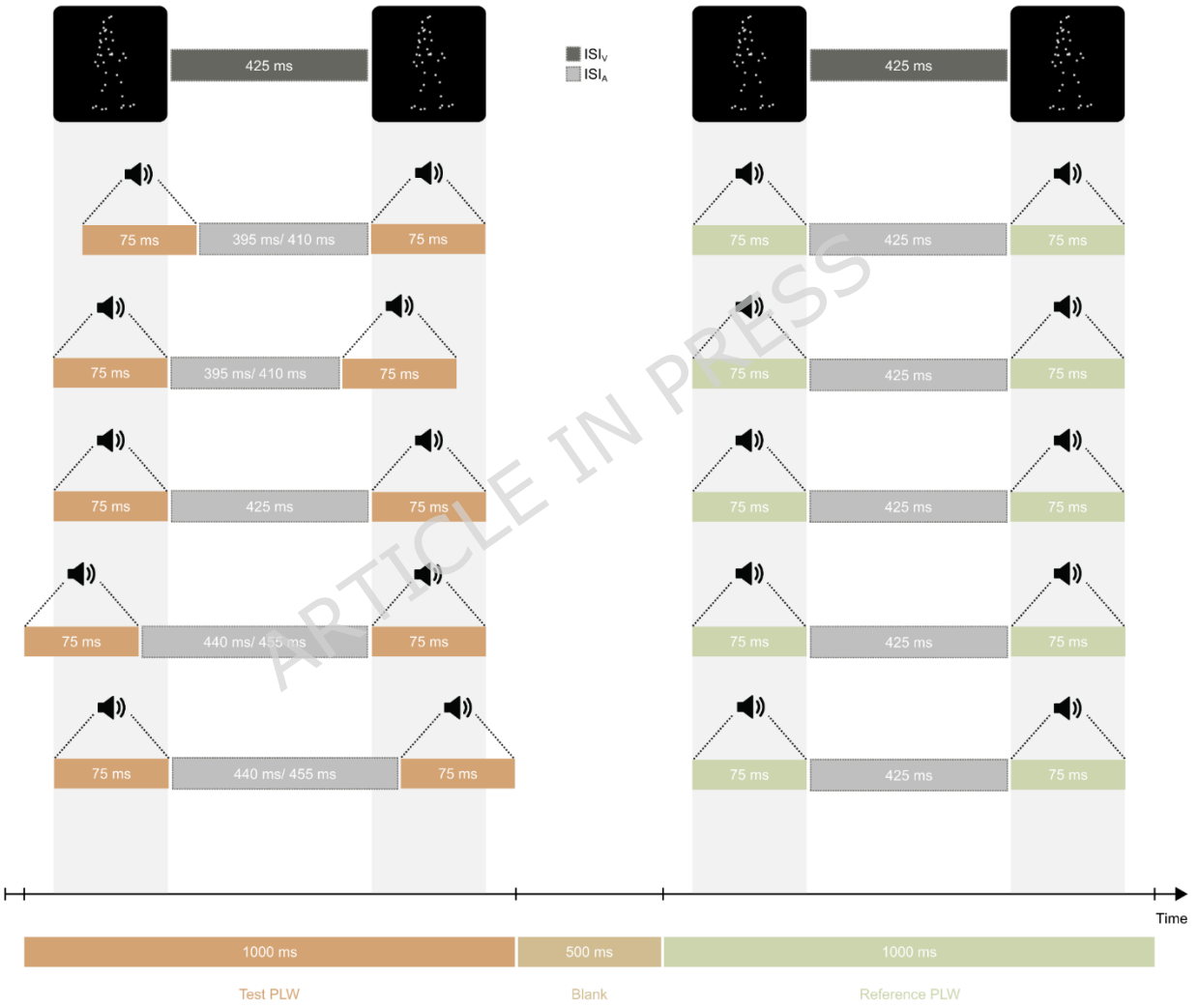


Figure 3. Schematic representation of experimental design and stimulus timing in Experiment 1b. The trial structure was identical to that described in Fig. 1. In the reference sequence, the visual footsteps and auditory signals were temporally aligned, with an ISI_V of 425 ms and a corresponding ISI_A of 425 ms. In the test sequence, the auditory intervals were either shorter or

ARTICLE IN PRESS

longer than the visual intervals (ISI_A = 395, 410, 440, or 455 ms), resulting in temporal shifts of ±15 ms or ±30 ms relative to the visual interval. To isolate the contribution of the first versus second footstep sound, either the first sound (onset condition) or the second sound (offset condition) was displaced. In the 425 ms ISI_A condition, the test and reference stimulus pairings were matched, providing a baseline without any temporal shift (ΔISI_A = 0 ms [test - reference]).

3.1.3. Data Analysis

The auditory timing factor (ISI_A) was converted into a signed continuous predictor representing the difference of auditory inter-step intervals between the test and the reference (ΔISI_A, in ms; negative values indicate a shorter auditory interval, and positive values indicate a longer auditory interval). ΔISI_A was centered at 0 ms (synchrony), so the intercept corresponds to the predicted log-odds at ΔISI_A = 0. The Onset/Offset factor indicated which sound (first vs. second) was temporally shifted. Because baseline trials (ΔISI_A = 0) lack a defined onset/offset designation, they were excluded from the analysis. A robustness check in which baseline trials were randomly assigned to onset and offset conditions in equal proportions is reported in the Supplementary Materials.

Following the same parsimony rationale as in Experiment 1a, we evaluated a mixed-effects specification with a random intercept at the by-participant level. The random-effect variance was estimated at ~ 0 (singular fit), and a GLMM-GLM comparison showed no improvement in fit (identical log-likelihood; $\Delta AIC \approx 2$, $\chi^2 = 0$, $p = 1$). Therefore, inference is equivalent to a fixed-effects logistic regression, and we report coefficients from the fitted generalized linear model. This model tested whether the likelihood of judging the test walker as faster varied linearly with ΔISI_A and whether this effect differed between onset and offset manipulations. Wald *z*-tests were used for inference. Planned contrasts were conducted using *emmeans*[45] with the Holm correction for multiple comparisons.

3.2. Results

ARTICLE IN PRESS

The generalized linear mixed model revealed a strong linear effect of auditory-visual asynchrony on perceived biological-motion speed (Fig. 4). Increasing the auditory inter-step interval (ΔISI_A) significantly reduced the likelihood that participants judged the test walker as faster (β = -0.0128, SE = 0.0015, z = -8.26, $p < 0.001$). This corresponds to an odds ratio of 0.82 per 15 ms increase (≈ 18% decrease in the odds of a "test faster" judgment for every 15 ms auditory delay). The intercept did not differ from zero ($\beta = 0.011$, $p = 0.76$), indicating no bias at synchrony ($\Delta ISI_A = 0$). Neither the main effect of OnsetOffset ($\beta = 0.017$, $p = 0.65$) nor its interaction with ΔISI_A ($\beta = 0.0018$, $p = 0.24$) reached significance, confirming that manipulating the first or second sound produced equivalent perceptual outcomes. Model-based predicted probabilities showed a clear monotonic pattern. Participants were most likely to report the test walker as faster when the auditory interval was 30 ms shorter than the visual interval ($P = 0.61$ [0.57-0.65]) and least likely when it was 30 ms longer ($P = 0.39$ [0.35-0.43]; Fig. 4).

Pairwise ΔISI comparisons within both onset and offset conditions revealed a highly consistent progression: every step toward a longer auditory interval (−30 → −15 → +15 → +30 ms) significantly decreased the odds of a "test faster" judgment (all zs > 5.0, all Holm-corrected ps < 0.0001). These results replicate and extend the findings of Experiment 1a, demonstrating that perceived biological-motion speed scales linearly with auditory-visual timing, independent of whether the first or second footstep sound is shifted. Increasing the auditory inter-step interval reliably decreased the probability of judging the test walker as faster, and this effect was comparable for onset- and offset-aligned sounds, indicating a consistent cross-modal temporal influence on perceived biological-motion speed. Thus, the cross-modal influence of auditory timing on biological-motion speed perception appears to reflect a temporally symmetric integration process over the footstep sequence, rather than a preferential weighting of later auditory events.

ARTICLE IN PRESS

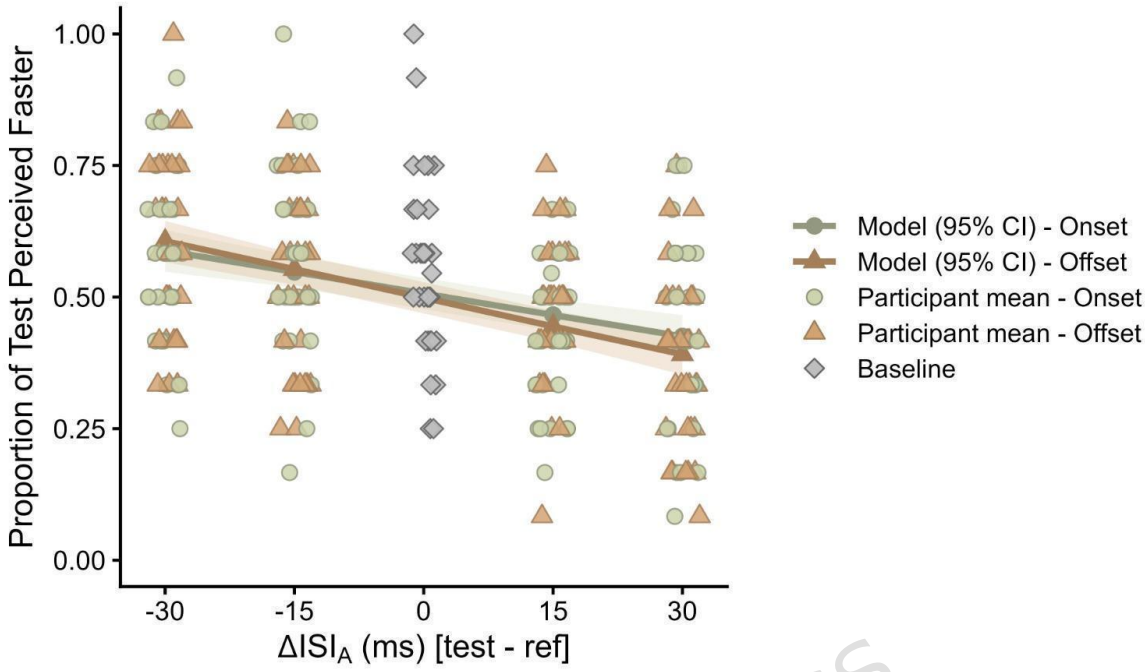


Figure 4. Proportion of "test faster" judgments as a function of auditory step-interval differences (ΔISI_A) in Experiment 1b. Each light green circle and light orange triangle represents a participant's mean proportion of "test faster" responses at each ΔISI_A condition. Individual means for the baseline condition (identical step intervals; ΔISI_A = 0 ms) are depicted by grey diamonds. Shorter auditory intervals for the test stimulus are denoted by negative ΔISI_A values, while longer ones are denoted by positive values. The dark green line and matching filled circles, as well as the dark orange line and matching colored filled triangles, demonstrate the population-level predicted probabilities from the binomial GLMM, with shaded areas indicating the model's 95% confidence interval. The GLMM analysis indicated a significant effect of ΔISI_A: a 15-ms increase in auditory interval decreased the likelihood of a "test faster" judgment by about 18%. Individual trajectories with connecting lines are presented separately in Supplementary Fig. S7 for onset (a) and offset (b) conditions.

4. Experiment 2a

Multisensory processing is shaped not only by temporal alignment but also by higher-level correspondences such as semantic congruency between sensory inputs. Previous research has demonstrated that when visual and auditory stimuli are semantically coherent—such as congruence between a voice and the perceived sex of the speaker—participants exhibit enhanced

ARTICLE IN PRESS

temporal discrimination. Specifically, Vatakis and Spence[48] found that temporally aligned and semantically matched stimuli led to stronger perceptual binding compared to mismatched pairs. Converging evidence from audiovisual speech perception and object recognition similarly indicated that semantic congruency facilitates cross-modal integration and perceptual clarity (e.g., Laurienti et al.[49]; Chen & Spence[50]). Building on this framework, and on evidence that biological motion engages specialized perceptual mechanisms sensitive to contextual cues[20,21], we examined whether the influence of auditory temporal structure on perceived speed of motion under audiovisual conditions depends on the semantic content of the sounds. Specifically, we paired point-light walkers with sounds that either plausibly corresponded to human locomotion (e.g., footsteps) or were semantically unrelated to walking (e.g., a coin sound). We tested the hypothesis that semantically matched sounds would exert a stronger modulation of speed judgments than unmatched sounds, reflecting an increased likelihood of cross-modal binding when auditory and visual signals convey compatible action information.

ARTICLE IN PRESS

4.1. Methods

4.1.1. Behavioral pilot study: Determination of sound content

We conducted a pilot study to characterize the semantic attributes of candidate sound stimuli to be used in the main experiment. Participants rated each sound on two dimensions, presented in a randomized order: (1) the degree to which the sound resembled footsteps, and (2) the likelihood that it was produced by a human. The study was programmed in PsychoPy (version 2024.2.4) using the Builder interface[51], which was also used to create response sliders. We used the same equipment and testing room as in the main and other experiments. Unlike the main experiment, this pilot study presented white sliders and white-colored questions (luminance = 0.21 cd/m²) against a mid-gray background (luminance = 16.21 cd/m²). Each trial

ARTICLE IN PRESS

presented a single sound (75 ms duration), followed by two rating questions displayed in random order on the screen. Responses were recorded via mouse clicks on five-point sliders (Supplementary Fig. S4). A session consisted of 110 trials, comprising 11 sounds presented 10 times each in randomized order.

Data were collected from twenty-six participants (18 females; age range: 18-29 years). The inclusion and exclusion criteria, as well as ethical procedures, were the same as those used in the main experiments. Statistical tests conducted on group-averaged values (Supplementary Figs. S5 and S6) confirmed that several sounds differed reliably from footstep sounds along the two semantic dimensions. Based on these results, five sound types were selected for the main experiment. The auditory stimuli consisted of footsteps, hand claps, water droplets, gamified coin, and jump sounds (hereinafter referred to as step, clap, drop, coin, and jump, respectively). This set included both sounds closely associated with human locomotion and sounds varying in perceived human origin and action relevance.

4.1.2. Participants, stimuli and procedure

Thirty-two adult volunteers took part in the experiment. The same participant cohort as in Experiments 1a and 1b was tested. Stimulus parameters, task structure, and procedural details followed those of Experiment 1a, with one critical modification: sound type was introduced as an additional experimental factor in the design. There were five sound conditions (step, clap, drop, coin, and jump), selected based on the outcomes of the pilot study. In addition to sound type, we included four main auditory time interval conditions and a baseline/synchronous condition (ISI_A = 365, 395, 455, 485 ms, with 425 ms as the baseline).

Each session comprised 25 unique conditions (5 sound types × 4 main ISI_A and a baseline condition for each sound type), with 10 repetitions per condition, totaling 250 trials per participant. To mitigate fatigue, given the

ARTICLE IN PRESS

increased trial count, mandatory break opportunities were provided every 50 trials. During these intervals, the fixation point changed color from white to red, signaling that participants could pause the task. Participants resumed the experiment at their own pace by pressing a key. All other aspects of the procedure, including practice blocks and inclusion/exclusion criteria, were identical to Experiment 1a.

4.1.3. Data Analysis

Auditory timing was modeled as a signed continuous predictor reflecting the difference in auditory inter-step intervals between test and reference stimuli (ΔISI_A). The intercept directly corresponds to the predicted log-odds at temporal synchrony ($\Delta ISI_A = 0$ ms). No mean-centering was applied, as ΔISI_A is already expressed relative to a meaningful zero point. Sound type was included as a categorical fixed effect indexing the semantic content of the accompanying sound.

The generalized linear mixed-effects model included ΔISI_A, sound type, and their interaction as fixed effects. Sound type comprised five levels (step, clap, drop, coin, jump), with the footstep condition specified as the reference level. Accordingly, coefficients for the remaining sound types reflect deviations from the footstep baseline. The model estimated whether the probability of judging the test stimulus as faster varied as a function of auditory timing and sound semantics. Random-effects structure and all other modeling choices were identical to those used in Experiments 1a and 1b.

4.2. Results

The proportion of trials in which the test PLW was judged as faster is shown in Figure 5. The GLMM analysis revealed that the auditory sound timing was the sole predictor of perceived speed. As ΔISI_A increased, the likelihood that participants judged test stimuli as faster decreased ($\beta = -0.01$, $SE = 0.001$, $z = -11.44$, $p < 0.001$). The intercept did not differ from zero ($\beta = 0.06$, $p =$

ARTICLE IN PRESS

0.26), indicating no systematic response bias at synchrony ($\Delta ISI_A = 0$). There was no evidence that the semantic content of sounds influenced perceived speed. Neither the main effects of sound type –coin ($\beta = 0.03$, $SE = 0.07$, $z = 0.45$, $p = 0.65$), drop ($\beta = 0.04$, $SE = 0.07$, $z = 0.56$, $p = 0.58$), jump ($\beta = 0.001$, $SE = 0.07$, $z = 0.02$, $p = 0.99$), or clap ($\beta = -0.04$, $SE = 0.07$, $z = 0.51$, $p = 0.61$)– nor their interactions with ΔISI_A reached significance (all ps ≥ 0.15). Notably, increasing auditory inter-step intervals reliably reduced the odds of a “test faster” judgment across all sound conditions (all zs > 9.93, all Holm-corrected ps < 0.0001).

ARTICLE IN PRESS

ARTICLE IN PRESS

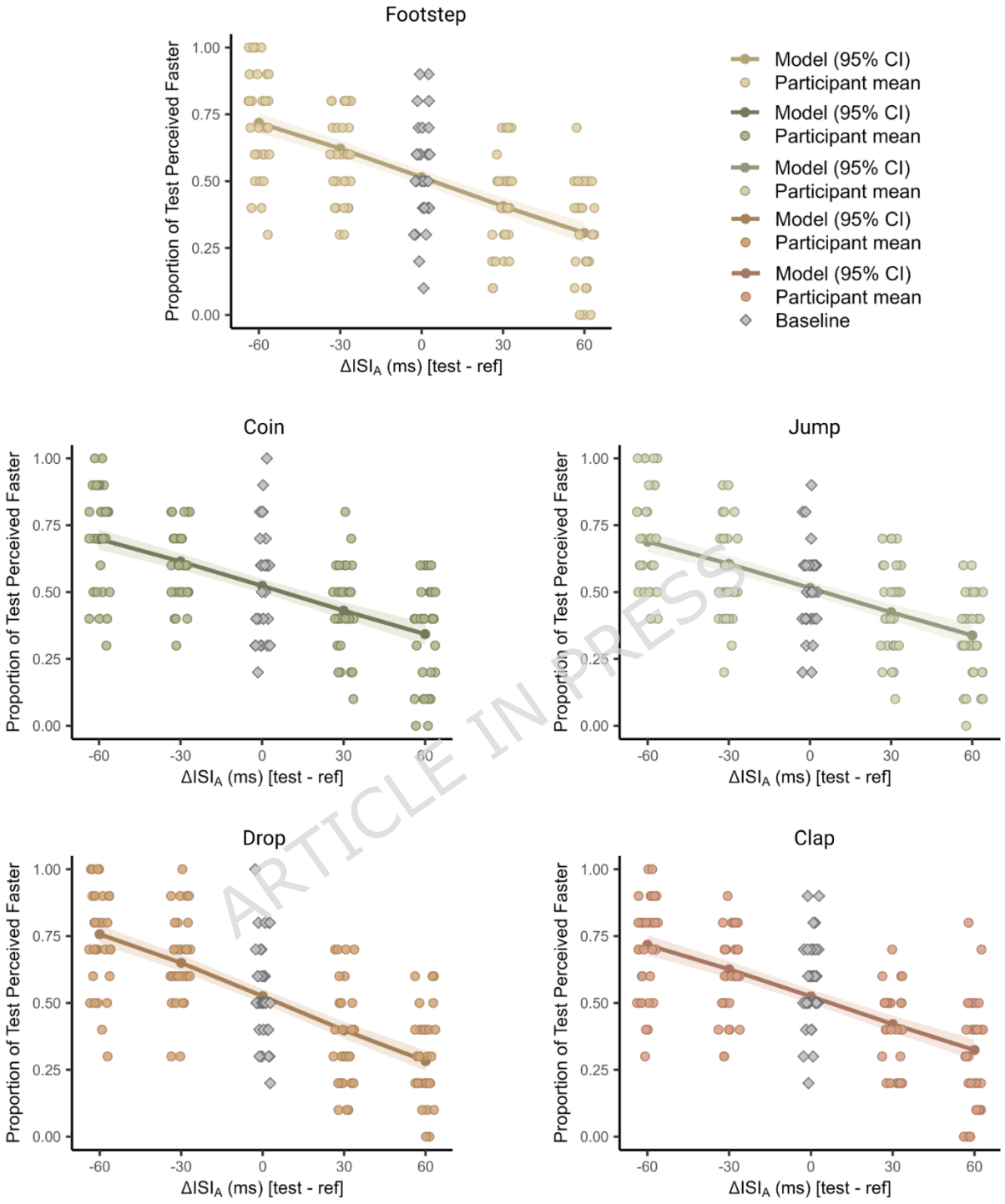


Figure 5. Proportion of "test faster" judgments as a function of auditory step-interval differences (ΔISI_A) in Experiment 2a. The results of five different sound type conditions are displayed in separate plots. In each plot, the colored lines with matching filled circles represent the population-level predicted probabilities from the binomial GLMM (shaded areas = 95% CI).

ARTICLE IN PRESS

Lighter-filled circles represent individual participants' mean proportion of "test faster" responses at each ΔISI_A condition, with gray diamonds depicting individual means for the baseline condition (identical step intervals; ΔISI_A = 0 ms). The negative and positive ΔISI_A values correspond to shorter (faster) and longer (slower) auditory intervals for the test walker, respectively.

5. Experiment 2b

Experiment 2a revealed no evidence that the semantic content of sounds modulated the influence of auditory timing on perceived biological motion speed. In that experiment, each sound had a duration of 75 ms, and the pilot study confirmed that participants could reliably distinguish sound categories within this time window. Nevertheless, it remained possible that the brief duration and limited exposure to individual sounds constrained the extent to which semantic information could influence audiovisual integration.

Experiment 2b was therefore designed to test whether increasing the perceptual salience of semantic cues would reveal a modulatory role of sound content. To this end, we extended the duration of auditory stimuli and increased the number of sound–step pairings within each trial. In addition, we enhanced the visual stimulus by increasing the number of gait cycles displayed. Together, these manipulations were intended to strengthen the availability of semantic information in both modalities and to assess whether greater semantic clarity alters the effect of auditory temporal structure on perceived speed of point-light walkers.

5.1. Methods

5.1.1. Participants

We conducted a separate priori power analysis using G*Power 3.1.9.7[29] to determine the required sample size for this experiment. Power calculations indicated that a minimum of 16 participants was sufficient to detect a medium effect size (f = 0.25) with an alpha of .05 and power of .95. To maintain partial consistency with the sample size used in Experiment 2a, which included 32

ARTICLE IN PRESS

participants, we recruited 25 volunteers (13 females; age range: 18-25 years). None of these participants had taken part in any of the previous experiments.

5.1.2. Stimuli and procedure

The basic parameters of the stimuli, speed discrimination paradigm, and procedure were the same as those described in Experiment 2a, with targeted modifications to the auditory and visual stimuli. Three types of sounds were selected to examine the influence of auditory semantic content on perceived speed of motion. Specifically, we included a footstep sound to represent a semantically congruent auditory stimulus, given its natural association with human locomotion depicted in the PLWs. To contrast this, a coin sound was introduced as a semantically incongruent stimulus, as it was demonstrated in our preliminary study to bear minimal resemblance to the act of walking. Additionally, a white noise was incorporated as a non-semantic control condition to assess the baseline influence of an acoustically complex but non-meaningful auditory input. This range of sound types allowed us to investigate how varying levels of semantic relevance modulate audiovisual interactions in biological motion perception. Critically, to increase auditory salience, we extended the onset and offset of the sounds by 10 ms, making them to last 95 ms. The duration of a single visual step was kept constant at 75 ms.

ARTICLE IN PRESS

The experiment consisted of two blocks. In the first block, PLWs completing one gait cycle were paired with extended auditory stimuli. In the second block, both auditory and visual stimulus durations were extended. Unlike previous experiments, PLWs were shown for 1.5 gait cycles (three steps over 1.5 seconds) instead of one cycle (two steps over 1 second). The sound aligned with the second step was always presented synchronously, while the timing of sounds for the first and third steps was systematically manipulated (Fig. 6).

ARTICLE IN PRESS

Each block included 135 trials: a factorial combination of 3 sound types and 4 main inter-stimulus intervals (ISI_A = 345, 375, 435, and 465 ms) with 10 repetitions (120 trials), supplemented by 5 repetitions of a baseline condition (ISI_A = 405 ms) for each sound type. Each 95-ms sound was presented at varying time intervals relative to the onset of each foot/visual step, while the interval between visual steps was held constant at 425 ms. For the reference stimulus, the auditory (ISI_A = 405 ms) and visual (ISI_V = 425 ms) time intervals were always fixed. As in previous experiments, the test stimuli varied across the five auditory time interval conditions.

ARTICLE IN PRESS

ARTICLE IN PRESS

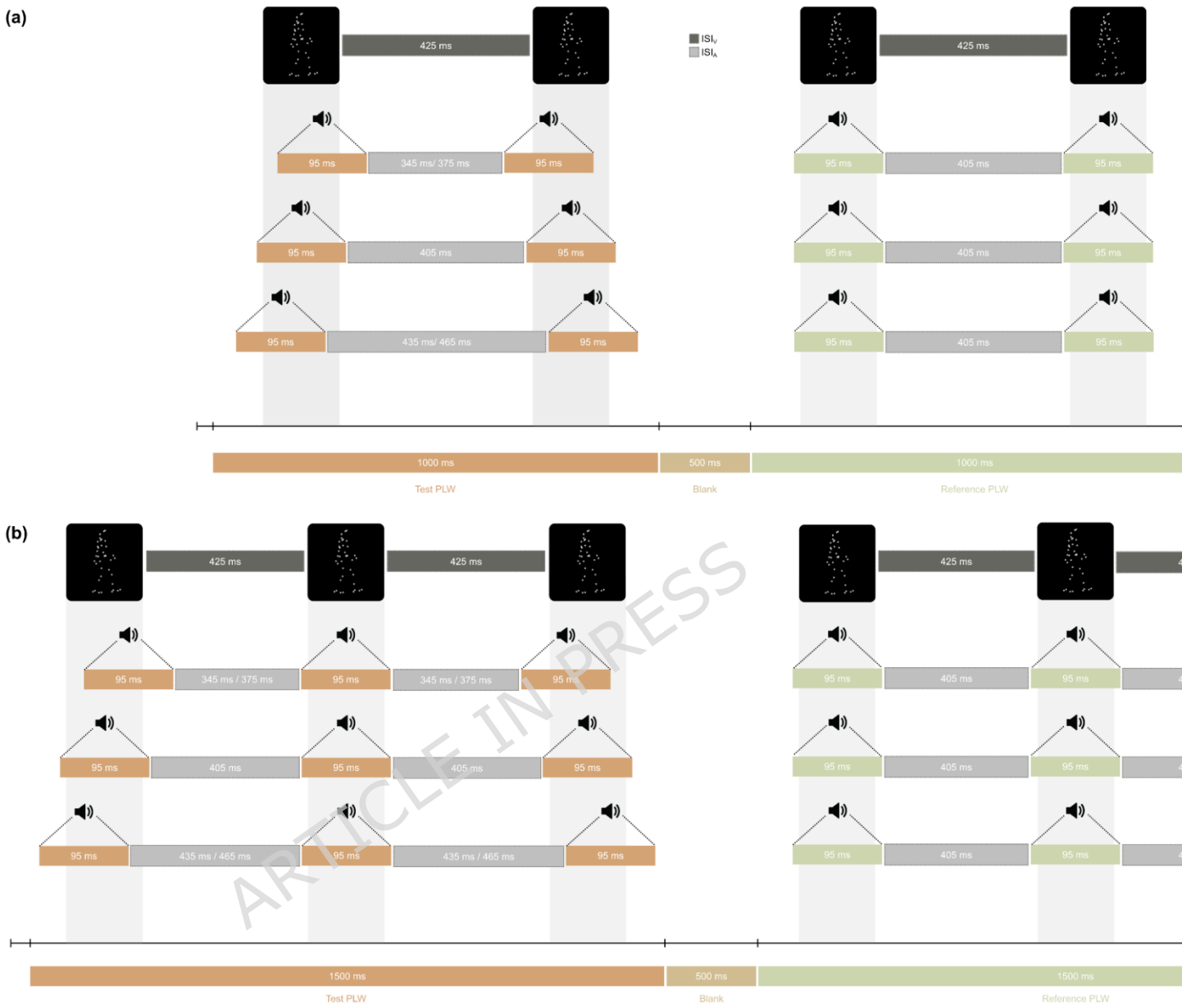


Figure 6. Schematic illustration of design and stimulus timing in Experiment 2b. **(a)** Each trial consisted of a test (left) and a reference (right) PLW sequence, each lasting 1000 ms and separated by a 500 ms blank interval. In the reference sequence, visual footsteps (75 ms duration) and sounds (95 ms duration) were aligned, with an ISI_V of 425 ms and an ISI_A of 405 ms. In the test sequence, auditory intervals were either shorter or longer than visual intervals (ISI_A = 345, 375, 435, or 465 ms), resulting in temporal shifts of ±30 ms or ±60 ms. **(b)** Each trial consisted of a test and reference PLW sequence lasting 1500 ms. This longer sequence resulted from the presentation of three steps instead of two. To maintain ±30 ms or ±60 ms temporal shifts, the second visual and auditory footsteps were aligned, and

ARTICLE IN PRESS

the auditory inter-step intervals for the first and third steps were manipulated accordingly (ISI_A = 345, 375, 435, or 465 ms).

5.1.3. Data Analysis

To effectively capture the distinct dynamics between the test and the reference stimuli, auditory timing was encoded as a signed continuous predictor, following the analytical approach established in earlier experiments. As in previous experiments, baseline trials (ΔISI_A = 0 ms), representing synchronous audiovisual pairings, were included as a level of the auditory timing predictor. The model also included sound type as a three-level (coin, step, and noise) categorical fixed factor, with the footstep sound designated as the reference level. Consequently, coefficients for the remaining sound types represent deviations from this baseline. All remaining aspects of model specification, estimation, and inference adhered to protocols established in previously reported experiments.

5.2. Results

The GLMM analysis demonstrated that auditory timing significantly modulates the perceived speed of biological motion (Fig. 7a). Specifically, as ΔISI_A increased, the probability of participants perceiving the test stimuli as 'faster' decreased (β = -0.01, SE = 0.001, z = -8.25, $p < 0.001$). At the synchronous condition (ΔISI_A = 0), no perceptual bias was observed, as the intercept did not statistically differ from zero ($\beta = 0.10$, $p = 0.12$). Regarding to semantic content of the auditory stimuli, results relative to the reference level—footstep—were mixed. The analysis yielded a significant main effect for coin sound (β = -0.25, SE = 0.09, z = -2.83, $p < 0.01$). In contrast, no significant main effect was observed for noise sound ($\beta = -0.01$, $SE = 0.09$, z = -0.16, p = 0.87). Furthermore, no significant interactions were found between sound type and ΔISI_A (coin x ΔISI_A: $\beta = 0.0001$, $SE = 0.002$, $z = 0.05$, $p = 0.96$; noise x ΔISI_A: $\beta = -0.001$, $SE = 0.002$, $z = -0.60$, $p = 0.55$). Despite the isolated significant main effect of coin sound , the overall pattern

ARTICLE IN PRESS

indicates no modulatory role of the semantic content of the sounds. In contrast, comparisons across auditory timing levels within each sound condition revealed a consistent temporal progression, with longer auditory intervals yielding fewer “test faster” responses (all *z*s > 8.18, all Holm-corrected *p*s < 0.0001).

Results from the second block are shown in Figure 7b and closely mirrored those of the first block. Auditory timing again strongly predicted perceived speed, with increasing ΔISI_A associated with a reduced likelihood of “test faster” judgments (β = -0.01, SE = 0.001, z = -8.23, p < 0.001). In the synchronous condition (ΔISI_A = 0), response patterns revealed no inherent bias, as the intercept was indistinguishable from zero (β = -0.04, p = 0.49). Notably, as in the first block, the semantic content of the paired sounds did not alter speed judgments of audiovisual stimuli. This independence is supported by the analysis of sound type, where neither the main effects for coin (β = 0.15, SE = 0.09, z = 1.69, p = 0.09), or noise (β = 0.09, SE = 0.09, z = 1.04, p = 0.30), nor their interaction with auditory timing (coin x ΔISI_A: β= 0.001, SE = 0.002, z = 0.67, p = 0.50; noise x ΔISI_A: β = 0.0002, SE = 0.002, z = 0.12, p = 0.91), yielded significant effects. Auditory timings for each sound type yielded significant results (all *z*s > 7.36, all Holm-corrected *p*s < 0.0001), with longer sound timings associated with fewer "test faster" responses, replicating the temporal pattern observed across earlier experiments.

ARTICLE IN PRESS

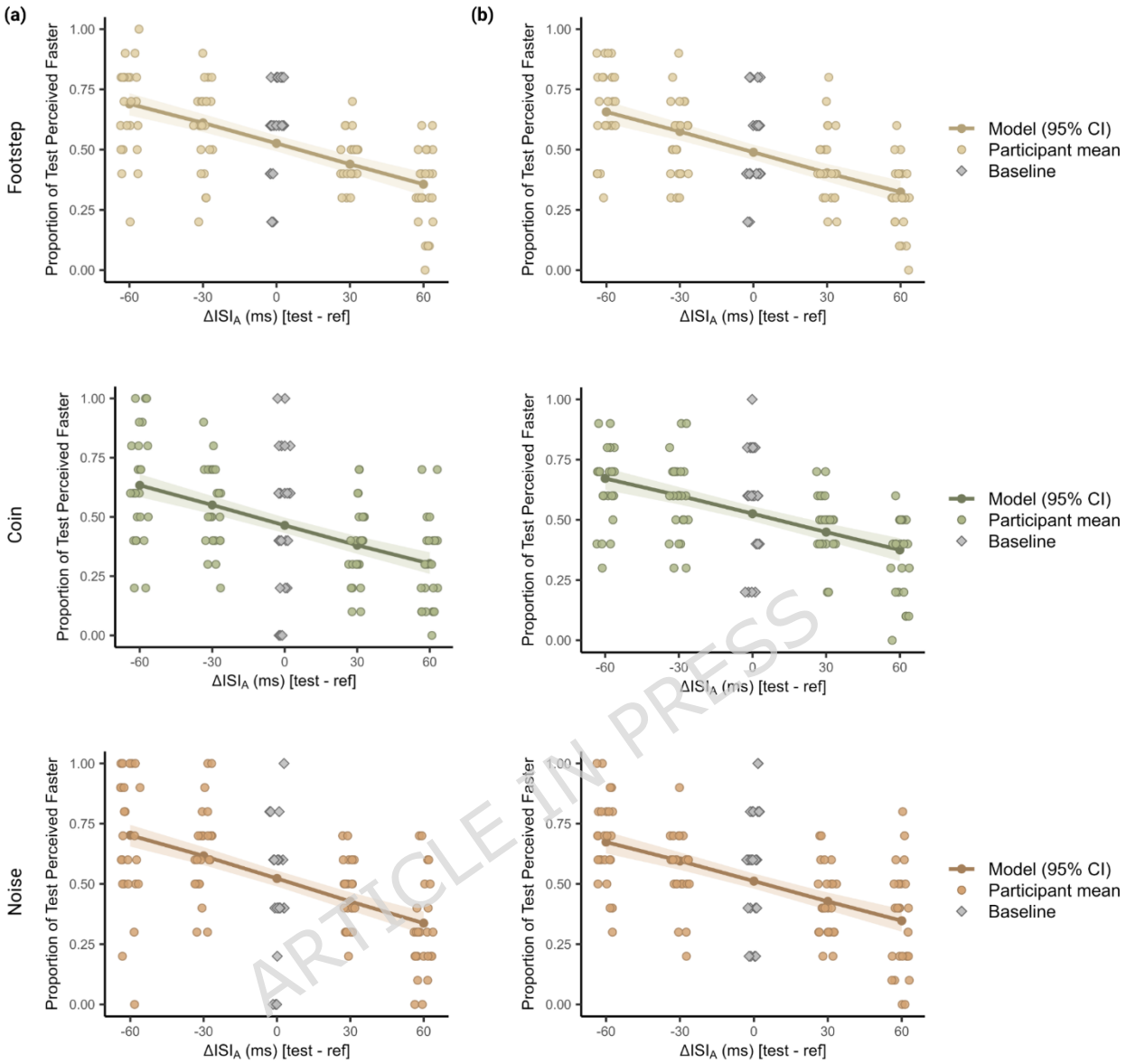


Figure 7. Proportion of "test faster" judgments as a function of auditory step-interval differences (ΔISI_A) in Experiment 2b. The results of Block 1 (two steps over 1 second) and Block 2 (three steps over 1.5 seconds) are shown on the left (a) and right (b), respectively. The proportion values of three different sound type conditions in each block are displayed in separate plots. The colored lines represent GLMM predicted probabilities (shaded areas = 95% CI). The lighter circles indicate individual participant means for 'test faster' responses, while grey diamonds indicate the baseline condition (ΔISI_A = 0 ms). The negative and positive ΔISI_A values correspond to shorter (faster) and longer (slower) auditory intervals for the test walker, respectively.

6. Discussion

ARTICLE IN PRESS

Across different experiments, we investigated whether the timing of auditory signals (e.g., footstep sounds) influences the perceived speed of biological motion. Consistent with prior work using simplified stimuli, our results showed that auditory temporal intervals significantly modulated perceived speed of motion even for a biologically meaningful, ecologically valid point-light walker, suggesting that auditory timing can influence visual temporal mechanisms in motion processing. Importantly, this effect persisted regardless of the semantic content of the accompanying sounds, reinforcing the notion that the influence of auditory timing on visual motion perception is largely automatic and bottom-up in nature. The observed insensitivity to sound semantics suggests that the temporal alignment between auditory and visual signals may be a dominant factor in crossmodal speed estimation. These findings not only extend the generalizability of temporal ventriloquism effects to biologically meaningful motion but also emphasize the importance of temporal structure over high-level semantic content in shaping perceptual outcomes.

ARTICLE IN PRESS

6.1. Semantic Congruency and Audiovisual Correspondence

Biological motion perception has frequently been examined in the context of audiovisual correspondence, where congruence between auditory and visual cues can facilitate detection and discrimination of motion. For example, step sounds temporally and semantically matched to the gait of a point-light walker have been shown to improve sensitivity to coherent motion[20-23]. Similarly, action-consistent sounds can bias perceived direction, modulate reaction times, and enhance the salience of a walking figure when the auditory and visual trajectories align[22,52,53]. These effects are typically interpreted as reflecting high-level crossmodal correspondence, where shared action-related meaning or spatiotemporal structure strengthens binding between modalities.

One of the central motivations of this work was to test whether semantic

ARTICLE IN PRESS

congruency or explicit action-related meaning amplifies auditory temporal effects on biological motion perception. Contrary to this possibility, neither the semantic category of the sound nor manipulations designed to enhance semantic salience reliably altered the magnitude or direction of the temporal bias. Even when auditory and visual elements were made more explicit and repeated, semantic enrichment did not modulate the effect. Importantly, this pattern does not necessarily conflict with Bayesian causal inference frameworks. Such accounts propose that observers infer whether signals originate from a common source by combining sensory evidence with prior expectations[54,55]. Semantic congruency may influence prior expectations regarding a common cause or stimulus properties such as speed. However, in the present paradigm, these priors may have carried relatively limited information, whereas the sensory signals provided strong and temporally precise evidence regarding event timing. Consequently, even if semantic context contributed to priors over common cause or expected speed, its influence on the final perceptual estimate may have been limited relative to the available temporal evidence. However, one apparent exception to this pattern requires careful examination. Although a main effect of sound type was observed for coin sound in the first block of Experiment 2b, several lines of evidence suggest this does not reflect a genuine perceptual phenomenon. First, this effect failed to replicate in the methodologically matched Experiment 2a and in the second block of Experiment 2b, despite comparable task demands and stimulus parameters. Second, the direction of the coin sound effect was inconsistent across experiments and blocks. Taken together, these considerations strongly suggest that the initial effect reflects statistical fluctuation rather than a stable perceptual bias, thereby reinforcing the robustness of our overall conclusions. Overall, the magnitude of temporal ventriloquism effects was mostly comparable for naturalistic step sounds and for other auditory stimuli lacking biological or action relevance. This negative finding is theoretically informative. It suggests that the auditory influence observed here is unlikely to be mediated by higher-level action understanding

ARTICLE IN PRESS

or conceptual matching between sound and motion. Instead, the data favor an interpretation in which temporal information operates at a relatively early or intermediate processing stage, shaping perceptual judgments through timing-based expectations rather than semantic inference. Within a causal inference framework, this pattern suggests a regime in which highly reliable temporal cues are weighted more strongly than semantic correspondence, leading to integration that is dominated by temporal structure despite variations in semantic congruency. In other words, while semantic matching can enhance the detectability and clarity of biological motion under noisy or ambiguous conditions, the modulation of speed perception by auditory time intervals appears to operate largely through automatic, bottom-up temporal interaction/binding mechanisms. Importantly, the results do not imply that semantic information is irrelevant for audiovisual perception more broadly. Rather, they suggest that, for speed judgments of biological motion, temporal alignment dominates over semantic correspondence. This distinction may help reconcile mixed findings in the literature regarding when and how biological relevance enhances multisensory processing, pointing to task demands and perceptual dimensions as critical moderators.

ARTICLE IN PRESS

Findings from previous research provide supporting evidence for this task and process dependency. Studies using paradigms that require temporal comparison often show that semantic congruency can facilitate multisensory processing, but this facilitation depends on maintaining temporally plausible audiovisual interactions.[24,52]. When temporal cues are sufficiently robust, as in the rhythmic gait of a point-light walker, semantic congruence appears to have little additional impact. By contrast, semantic information can become influential when temporal cues are weak, ambiguous, or degraded. For instance, Thomas and Shiffrar[21] demonstrated that semantic matching alone can facilitate the detection of biological actions even without precise temporal synchrony, whereas other studies have shown that semantic facilitation diminishes or disappears when synchrony is disrupted[24,56]. In the

ARTICLE IN PRESS

context of our paradigm, both congruent footstep sounds and incongruent sounds, such as coins, may nevertheless provide reliable temporal information. The constant presence of this temporal structure likely reduced the need for semantic coherence, since speed estimation depends strongly on temporal information and processing. Understanding how these mechanisms interact will require future work systematically varying the reliability of both semantic congruence and temporal alignment, ideally with ecologically valid motion and sound combinations.

6.2. Implications for Audiovisual Interactions and Temporal Binding

A notable aspect of the present findings is the insensitivity of the observed temporal ventriloquism effects to manipulations of sound onset and offset. This is consistent with previous evidence that the temporal binding window (TBW) for biological motion is broader than that for non-biological motion[24,25]. A broader TBW implies that the perceptual system tolerates greater temporal misalignments between auditory and visual events while still binding them into a unified percept. For BM, which is characterized by complex yet highly predictable kinematic patterns, such tolerance likely reflects adaptive tuning: the visual system can accommodate moderate temporal discrepancies without disrupting the perception of coherent human movement, maintaining functional integration across modalities even when auditory timing is perturbed. In addition to this temporal tolerance, we also found that the effects of auditory timing on perceived BM speed persisted regardless of the semantic content of the sound. As discussed earlier, this insensitivity to semantic congruency aligns with prior work showing that temporal ventriloquism and related audiovisual timing effects are not strictly dependent on feature similarity or crossmodal correspondence[57,58]. Within the context of BM, this suggests that temporal binding is primarily driven by the overall/general temporal and structural relationships between signals rather than by higher-level semantic associations. Notably, previous studies using simpler audiovisual stimuli have often reported asymmetric temporal

ARTICLE IN PRESS

binding windows, with differential sensitivity to auditory-leading and visual-leading offsets[59-63]. In contrast, the present findings did not reveal clear onset–offset asymmetries. This apparent discrepancy may reflect differences in stimulus complexity, task demands, and the relative reliability of temporal cues, as discussed in the following parts of this section.

From a theoretical standpoint, these findings are compatible with Bayesian causal inference models of multisensory integration[54,55]. In such models, the brain evaluates the probability that auditory and visual signals arise from a common source, weighting available cues according to their reliability. For BM, rich and ecologically valid kinematic patterns may act as strong priors for a common cause, thereby increasing tolerance for featural or semantic mismatches while maintaining sensitivity to temporal structure at some level. This interpretation aligns with Chen et al.'s[28] partial-cue integration account, in which overall temporal structure and motion dynamics can receive greater weighting than featural correspondence. Overall, the persistence of auditory timing effects in BM under both onset–offset and semantic manipulations suggests that temporal ventriloquism in BM may inherently include basic principles of audiovisual binding mechanisms while maintaining a specialized and ecologically tuned integration system. This system may prioritize temporal cues from multiple modalities, maintain robust binding despite moderate temporal disparities, and flexibly adjust weighting according to motion structure and ecological relevance. Such flexibility ensures that perceptual coherence is preserved for socially and behaviorally important signals, even under suboptimal temporal or featural conditions.

In the present design, we manipulated the timing of sounds relative to the visual footsteps, yet our conditions may not be sufficient to fully characterize asymmetries in onset–offset manipulations within the temporal binding window, particularly given that the temporal offsets used here (15–30 ms) likely fall within this window and may have been perceived as effectively synchronous. Although temporal sensitivity for simple audiovisual stimuli can

ARTICLE IN PRESS

reach approximately 25–50 ms[60,62,63], the use of more complex and naturalistic stimuli may broaden the effective binding window. The current findings nevertheless suggest that biological motion integrates auditory timing over a broader temporal range. A more direct assessment is required to confirm this property in the context of temporal ventriloquism. As Chen et al.[28] demonstrated, dissociating onset and offset manipulations can reveal the relative weighting of these temporal cues, offering deeper insight into the mechanisms underlying temporal ventriloquism. Moreover, their partial-cue integration framework suggests that overall temporal structure and local (micro) timing cues may contribute differently to audiovisual binding. Future work focused on mapping the binding window for BM using fine-grained manipulations of auditory onset and offset, and employing larger temporal discrepancies (e.g., ~100 ms), will be informative to explicitly determine whether the observed effects are driven primarily by the overall temporal structure or by micro-level timing perturbations. Such an approach would allow for disentangling the relative contributions of global temporal regularity and local timing shifts, under conditions in which audiovisual signals are more clearly perceived as asynchronous.

6.3. Putative Neural Substrates

Auditory contributions to the perception of BM speed remain poorly understood, yet available evidence suggests an overlap between cortical areas supporting BM and the parietal regions identified in source localization analyses of audiovisual apparent motion[15]. Neuroimaging and electrophysiological studies have shown that BM perception engages a distributed network including the superior temporal sulcus (STS), premotor cortex, and inferior parietal lobule[64-66]. Critically, the STS appears to be involved in integrating visual motion signals with ecologically relevant sounds, showing sensitivity to both temporal alignment of cross-modal inputs and their semantic congruence[59,67-69].

ARTICLE IN PRESS

Our finding—that auditory time intervals modulate perceived BM speed independently of semantic congruence—is consistent with a hierarchical integration framework in which early interactions in the temporal domain, mediated by low- and mid-level sensory areas, feeds into higher-level BM-selective networks. Prior EEG work on audiovisual apparent motion has shown that auditory timing modulates neural activity at both early and late stages of motion processing[15,17], with effects distributed over medial parieto-occipital, parietal, right centro-parietal, and frontal scalp sites. Source localization further implicates right parietal, superior and transverse temporal, and insular cortices, with later effects extending into right fronto-temporal regions including the superior frontal gyrus. Within the context of BM, such a dual-stage framework—early sensory capture followed by higher-order interpretative processing—could account for the persistence of temporal ventriloquism effects despite semantic mismatches. This broad and flexible integration mechanism would ensure robustness of motion perception in naturalistic environments where sound identity, location, and meaning may vary, but temporal coordination remains a reliable cue to common cause.

ARTICLE IN PRESS

However, these conclusions should be considered in light of several methodological considerations. First, while the auditory manipulations robustly affected perceived speed, the study was not designed to pinpoint the precise processing stage at which auditory timing exerts its influence. The data, therefore, cannot distinguish conclusively between early sensory interactions and later decisional or criterion-setting mechanisms. Second, although semantic salience was increased in later experiments, semantic processing itself was not independently measured. Thus, null effects should be interpreted as evidence of limited impact on speed judgments rather than the absence of semantic processing per se.

7. Conclusion

ARTICLE IN PRESS

To conclude, our findings demonstrate that auditory timing reliably modulates the perceived speed of biological motion. This influence persisted regardless of the semantic content of sounds or variations in their onset and offset timing, suggesting a robust temporal interaction/binding mechanism between vision and hearing. These findings refine existing accounts of multisensory processing by suggesting that biological relevance does not necessarily confer semantic dominance and that temporal structure can act as a key organizing principle for perceptual judgments. Results of the study highlight the fundamental role of multisensory temporal processing in shaping everyday perception, enabling us to interpret and respond to dynamic events such as walking footsteps in a crowd.

Funding This work was supported by the Scientific and Technological Research Council of Türkiye (BIDEB 2211-A) and the Turkish Academy of Sciences (TUBA-GEBIP Award).

Author contributions S.O., S.K.Y., and H.K. conceived of the study. S.O. collected the data. S.O. and S.K.Y. constructed and implemented the data analysis pipeline, and S.O. ran all statistical analyses. S.O., S.K.Y., and H.K. interpreted the results. S.O. wrote the first draft of the manuscript. S.O., S.K.Y., and H.K. edited the manuscript and approved the final version.

Data and code availability The dataset and experimental materials of the current study are available from the corresponding author on request. Any access to the data will be granted in accordance with the informed consent signed by the participants. The custom analysis code used in this study is publicly available at https://doi.org/10.5281/zenodo.20595175.

Declarations

Conflicts of interest The authors have no actual or potential conflicts of interest.

Ethical approval All experimental procedures were in accordance with the

ARTICLE IN PRESS

Declaration of Helsinki and international guidelines and approved by the local ethics committee.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Soto-Faraco, S., Kingstone, A., & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia*, *41*(13), 1847–1862. https://doi.org/10.1016/s0028-3932(03)00185-4

2. Soto-Faraco, S., & Väljamäe, A. (2012). Multisensory interactions during motion perception: From basic principles to media applications. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. CRC Press/Taylor & Francis.

3. Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception & Psychophysics*, *75*(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

4. Nishida, S. (2011). Advancement of motion psychophysics: Review 2001-2010. *Journal of Vision*, *11*(5), 11. https://doi.org/10.1167/11.5.11

5. Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current biology, 18*(16), 1262–1266. https://doi.org/10.1016/j.cub.2008.07.066

6. Kafaligonul, H., & Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals. *Journal of Vision*, *10*(12), 31. https://doi.org/10.1167/10.12.31

7. Ogulmus, C., Karacaoglu, M., & Kafaligonul, H. (2018). Temporal ventriloquism along the path of apparent motion: Speed perception

ARTICLE IN PRESS

under different spatial grouping principles. *Experimental Brain Research, 236*(3), 629-643. https://doi.org/10.1007/s00221-017-5159-1

8. Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception, 36*(7), 1089–1103. https://doi.org/10.1068/p5741

9. Kafaligonul, H., & Stoner, G. R. (2012). Static sound timing alters sensitivity to low-level visual motion. *Journal of Vision*, *12*(11), 2. https://doi.org/10.1167/12.11.2

10. Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval. *Experimental Brain Research, 203*(4), 723–735. https://doi.org/10.1007/s00221-010-2286-3

11. Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*(4), 718–725. https://doi.org/10.3758/BF03194432

12. Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*(1), 154–163. https://doi.org/10.1016/S0926-6410(03)00089-2

13. Duyar, A., Pavan, A., & Kafaligonul, H. (2022). Attentional modulations of audiovisual interactions in apparent motion: Temporal ventriloquism effects on perceived visual speed. *Attention, Perception & Psychophysics*, *84*(7), 2167–2185. https://doi.org/10.3758/s13414-022-02555-7

14. Yilmaz, S. K., & Kafaligonul, H. (2024). Attentional demands in the visual field modulate audiovisual interactions in the temporal

ARTICLE IN PRESS

domain. *Human Brain Mapping*, *45*(12), e70009. https://doi.org/10.1002/hbm.70009

15. Kaya, U., & Kafaligonul, H. (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

16. Akdogan, I., Aydin, S., & Kafaligonul, H. (2025). Dynamic reorganization of functional networks underlying audiovisual interactions. *Scientific Reports*, *15*(1), 39658. https://doi.org/10.1038/s41598-025-23390-4

17. Kaya, U., Yildirim, F. Z., & Kafaligonul, H. (2017). The involvement of centralized and distributed processes in sub-second time interval adaptation: An ERP investigation of apparent motion. *The European Journal of Neuroscience*, *46*(8), 2325–2338. https://doi.org/10.1111/ejn.13691

18. Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201–211. https://doi.org/10.3758/BF03212378

19. Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73. https://doi.org/10.1146/annurev.psych.57.102904.190152

20. Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of Vision*, *10*(12), 14. https://doi.org/10.1167/10.12.14

ARTICLE IN PRESS

ARTICLE IN PRESS

21. Thomas, J. P., & Shiffrar, M. (2013). Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony. *Journal of Vision, 13*(14), Article 8. https://doi.org/10.1167/13.14.8

22. Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia*, *45*(3), 523–530. https://doi.org/10.1016/j.neuropsychologia.2005.12.012

23. Wuerger, S. M., Parkes, L., Lewis, P. A., Crocker-Buque, A., Rutschmann, R., & Meyer, G. F. (2012). Premotor cortex is sensitive to auditory-visual congruence for biological motion. *Journal of Cognitive Neuroscience*, *24*(3), 575–587. https://doi.org/10.1162/jocn_a_00173

24. Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: Judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological Science*, *19*(5), 469–475. https://doi.org/10.1111/j.1467-9280.2008.02111.x

25. Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, *6*(3), 260–268. https://doi.org/10.1167/6.3.6

26. Ioannucci, S., & Vetter, P. (2025). Semantic audio-visual congruence modulates visual sensitivity to biological motion across awareness levels. *Cognition*, *262*, 106181. https://doi.org/10.1016/j.cognition.2025.106181

27. Boyce, W. P., Whiteford, S., Curran, W., Freegard, G., & Weidemann, C. T. (2020). Splitting time: Sound-induced illusory visual temporal fission and fusion. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(2), 172–201. https://doi.org/10.1037/xhp0000703

ARTICLE IN PRESS

ARTICLE IN PRESS

28. Chen, L., Zhou, X., Müller, H. J., & Shi, Z. (2018). What you see depends on what you hear: Temporal averaging and crossmodal integration. *Journal of Experimental Psychology: General, 147*(12), 1851–1864. https://doi.org/10.1037/xge0000487

29. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

30. World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

31. Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436. https://doi.org/10.1163/156856897x00357

32. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442. https://doi.org/10.1163/156856897x00366

33. Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement), 14.

34. Hodgins, J. (2015). *CMU graphics lab motion capture database*. Carnegie Mellon University. http://mocap.cs.cmu.edu/

35. Biomechanical Toolkit. (2018). *Mokka: Motion Kinematic and Kinetic Analyser* [Computer software]. https://biomechanical-toolkit.github.io/mokka/

36. Progin, P., Faivre, N., Brooks, A., Chang, W., Mercier, M., Schwabe, L., Do, K. Q., & Blanke, O. (2020). Somatosensory-visual

ARTICLE IN PRESS

effects in visual biological motion perception. *PLOS ONE*, *15*(6), e0234026. https://doi.org/10.1371/journal.pone.0234026

37. Saygin, A. P., Wilson, S. M., Hagler, D. J., Jr., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *The Journal of Neuroscience, 24*(27), 6181–6188. https://doi.org/10.1523/JNEUROSCI.0504-04.2004

38. Shen, L., Lu, X., Yuan, X., Hu, R., Wang, Y., & Jiang, Y. (2023). Cortical encoding of rhythmic kinematic structures in biological motion. *NeuroImage*, *268*, 119893. https://doi.org/10.1016/j.neuroimage.2023.119893

39. Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11656–11661. https://doi.org/10.1073/pnas.191374198

40. Audacity Team. (2025). *Audacity* (Version 3.4.2) [Computer software]. https://www.audacityteam.org/

41. Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

42. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

43. R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

ARTICLE IN PRESS

44. Posit Team. (2024). *RStudio: Integrated development environment for R*. Posit Software, PBC. http://www.posit.co/

45. Lenth, R. V. (2024). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.10.0) [R package]. https://CRAN.R-project.org/package=emmeans

46. Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

47. Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception & Psychophysics*, *72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

48. Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics, 69*(5), 744–756. https://doi.org/10.3758/BF03193776

49. Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405–414. https://doi.org/10.1007/s00221-004-1913-2

50. Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. https://doi.org/10.1016/j.cognition.2009.10.012

51. Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2:

ARTICLE IN PRESS

Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

52. Mendonça, C., Santos, J. A., & López-Moliner, J. (2011). The benefit of multisensory integration with biological motion signals. *Experimental Brain Research*, *213*(2-3), 185–192. https://doi.org/10.1007/s00221-011-2620-4

53. Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PLOS ONE*, *6*(2), e14725. https://doi.org/10.1371/journal.pone.0014725

54. Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE, 2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

55. Shams, L. (2012). Early integration and Bayesian causal inference in multisensory perception. In M. M. Murray & M. T. Wallace (Eds.), *Frontiers in the Neural Bases of Multisensory Processes* (pp. 217-231). CRC Press.

56. Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, *9*(4), 1–7. https://doi.org/10.1167/9.4.25

57. Boyce, W.P., Lindsay, A., Zgonnikov, A., Rañó, I., & Wong-Lin, K. (2020). Optimality and limitations of audio-visual integration for cognitive systems. *Frontiers in Robotics and AI, 7*, 94. https://doi.org/10.3389/frobt.2020.00094

58. Klimova, M., Nishida, S., & Roseboom, W. (2017). Grouping by feature of cross-modal flankers in temporal ventriloquism. *Scientific Reports, 7,* 7615. https://doi.org/10.1038/s41598-017-06550-z

ARTICLE IN PRESS

59. Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex, 18*(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187

60. Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental brain research, 167*(4), 635–640. https://doi.org/10.1007/s00221-005-0067-1

61. Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*(3), 468–478. https://doi.org/10.1016/S0926-6410(03)00074-0

62. Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*(2), 198–210. https://doi.org/10.1007/s00221-003-1536-z

63. Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics, 67*(3), 531–544. https://doi.org/10.3758/BF03193329

64. Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, *12*(5), 711–720. https://doi.org/10.1162/089892900562417

65. Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 435–445. https://doi.org/10.1098/rstb.2002.1221

ARTICLE IN PRESS

ARTICLE IN PRESS

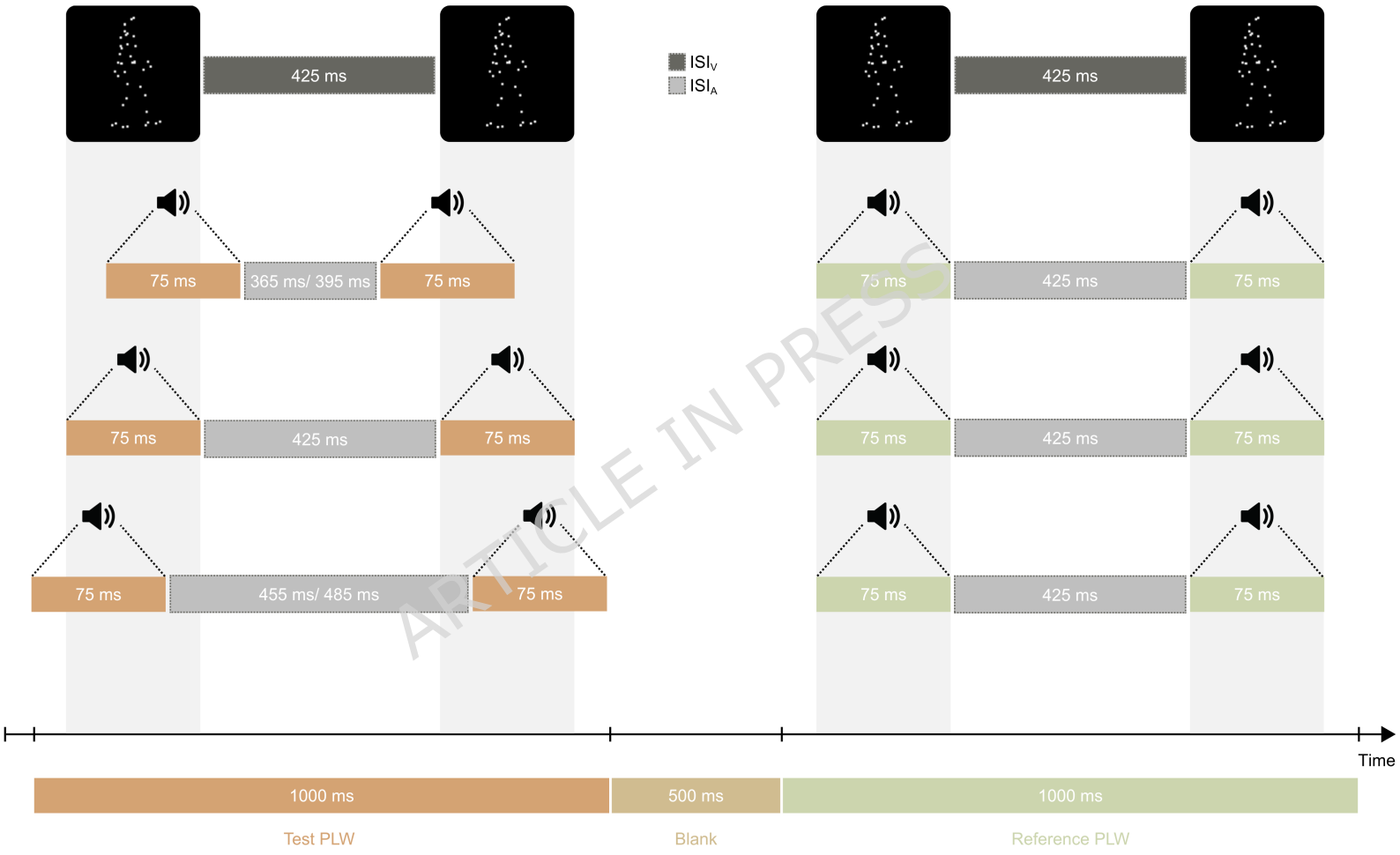
66. Saygin A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, *130*(Pt 9), 2452–2461. https://doi.org/10.1093/brain/awm162

67. Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–823. https://doi.org/10.1016/s0896-6273(04)00070-4

68. van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, *17*(4), 962–974. https://doi.org/10.1093/cercor/bhl007

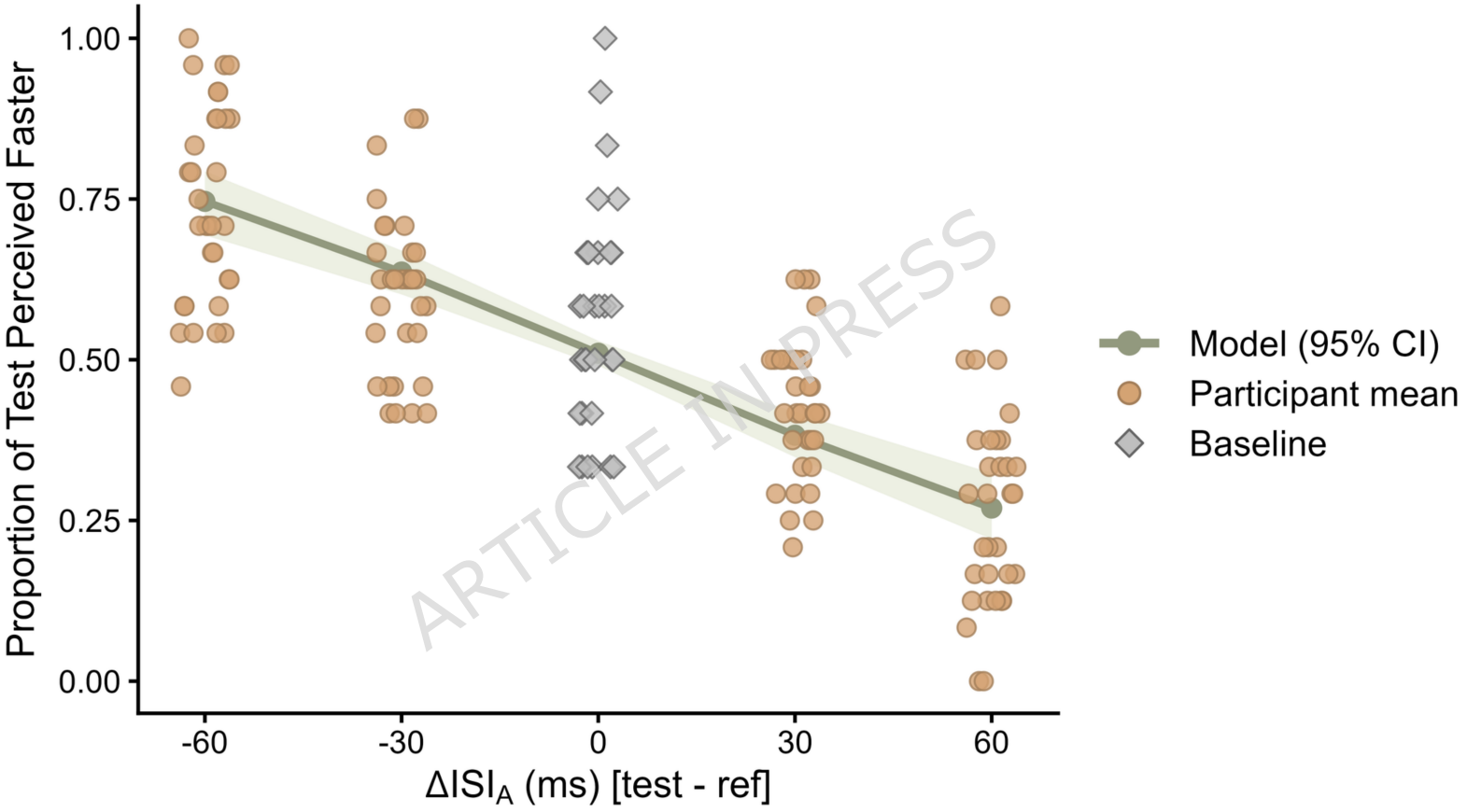
69. Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*, *30*(7), 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010

ARTICLE IN PRESS

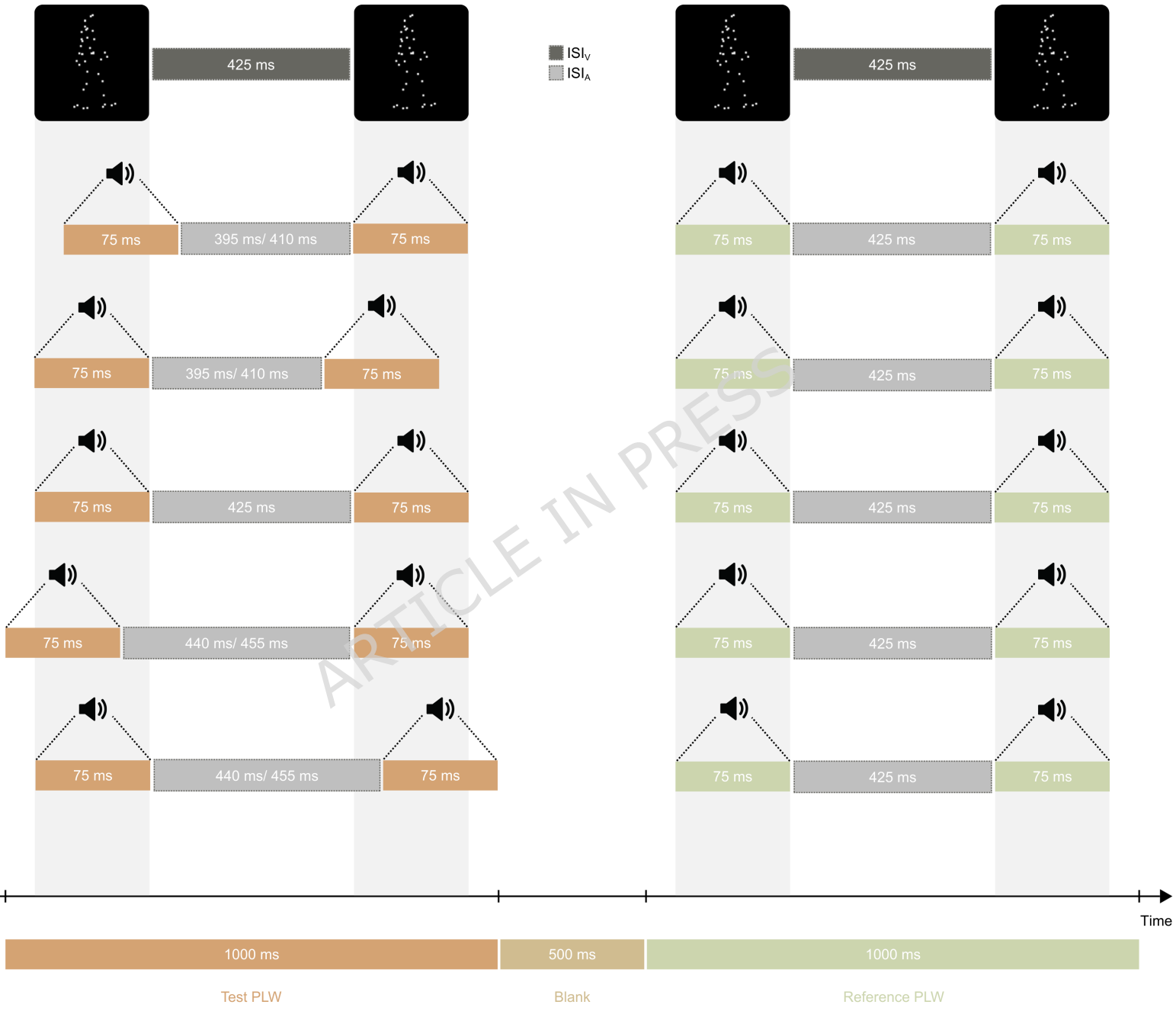


ARTICLE IN PRESS

ARTICLE IN PRESS

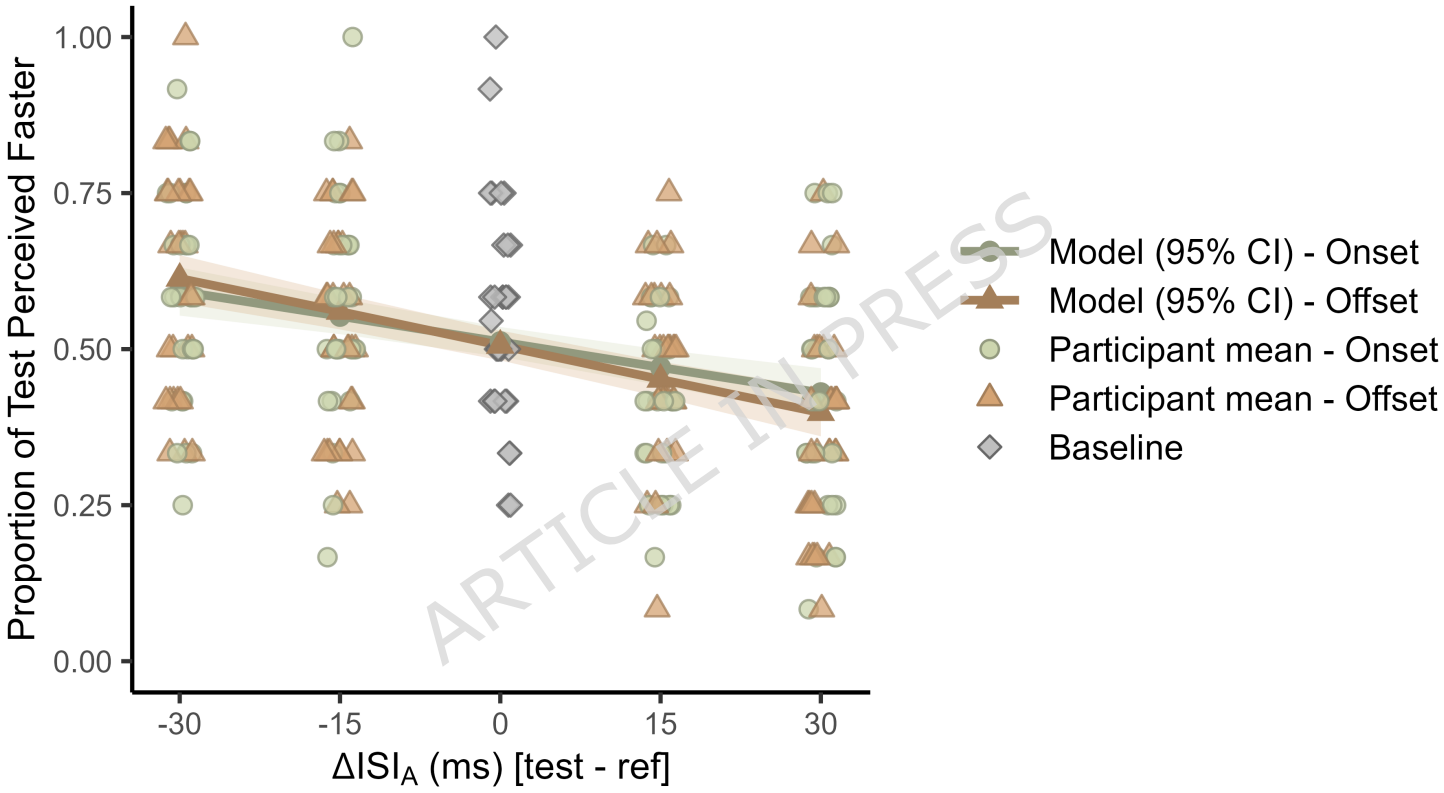


ARTICLE IN PRESS



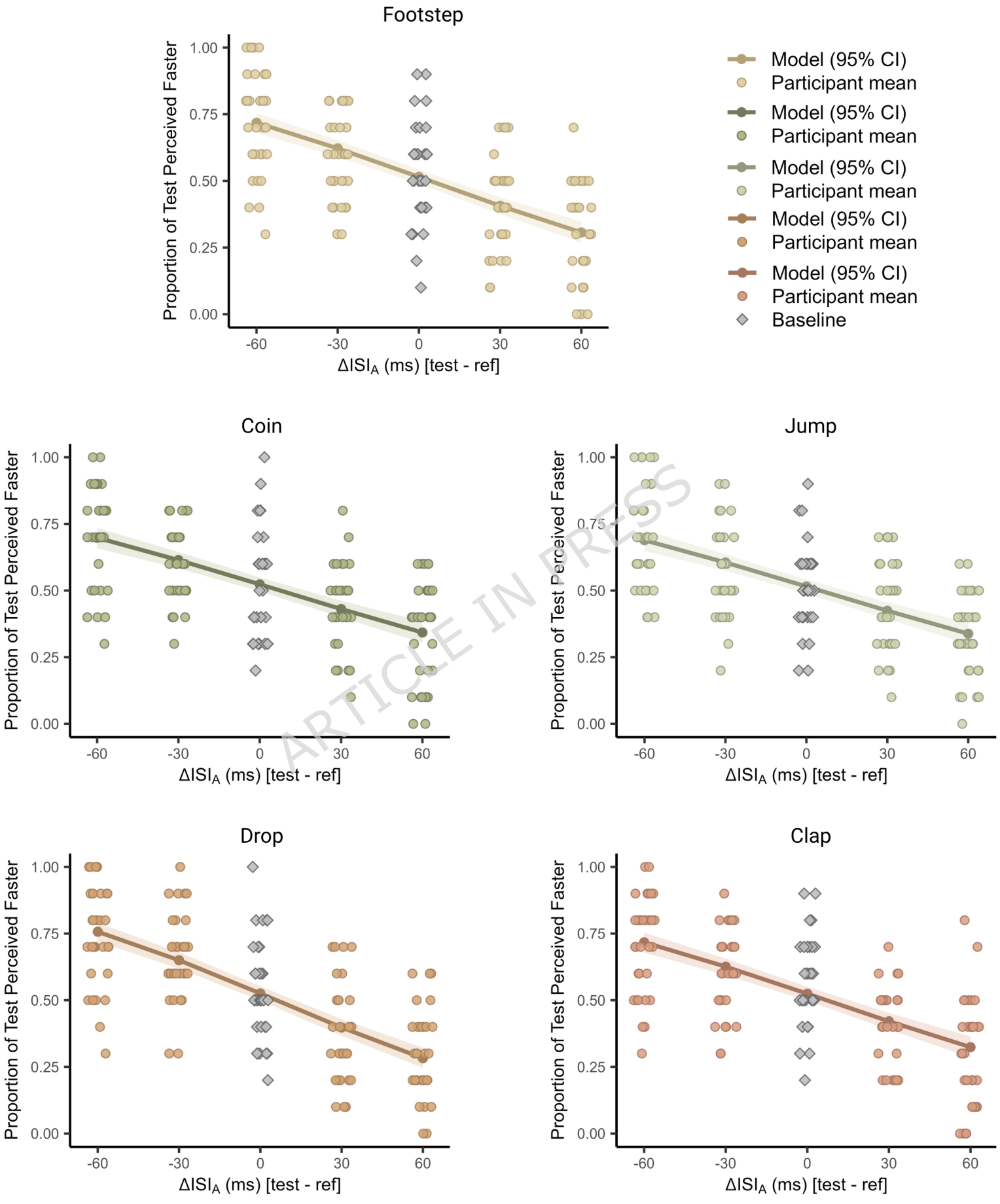
ARTICLE IN PRESS

ARTICLE IN PRESS



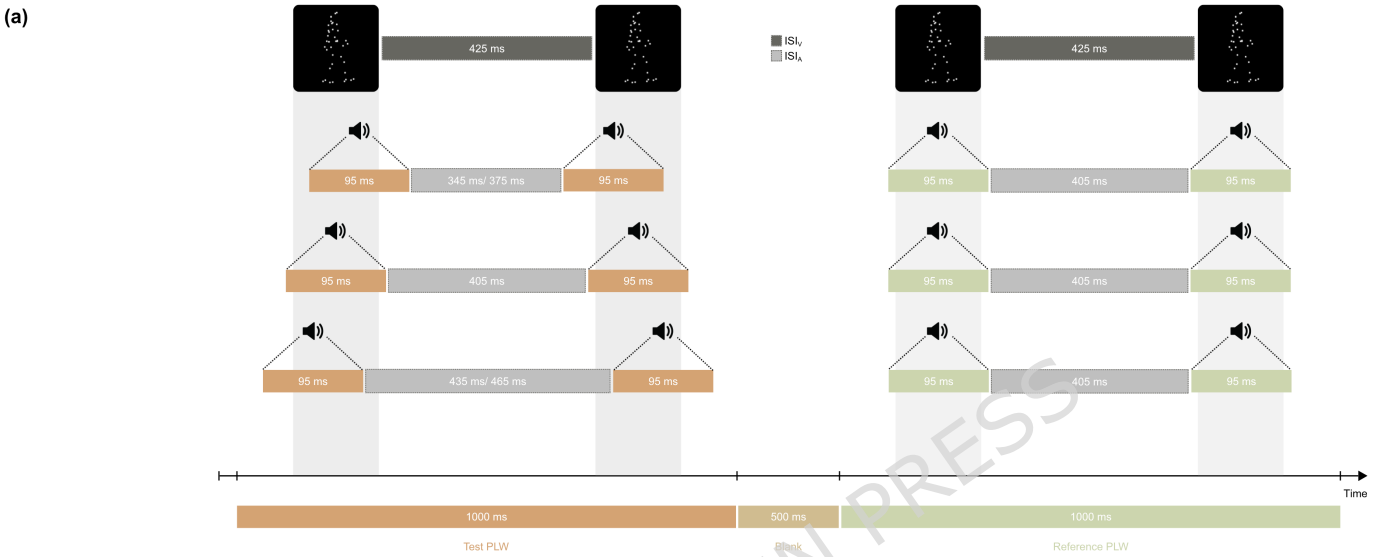
ARTICLE IN PRESS

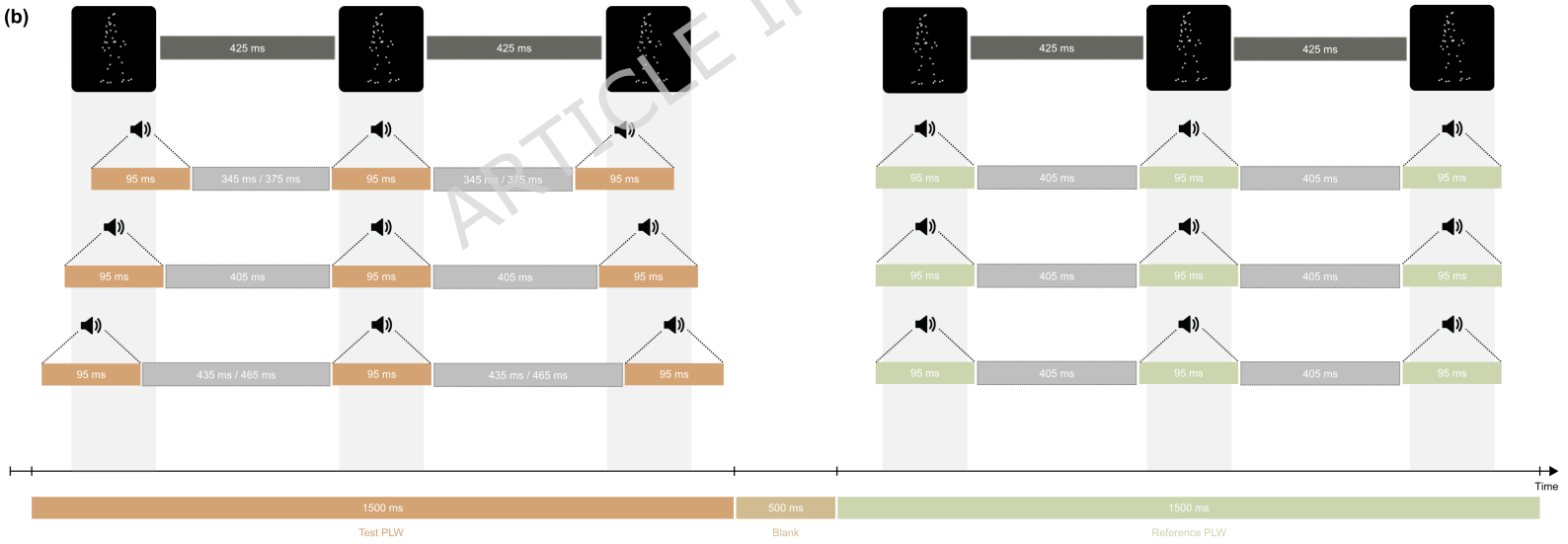
ARTICLE IN PRESS



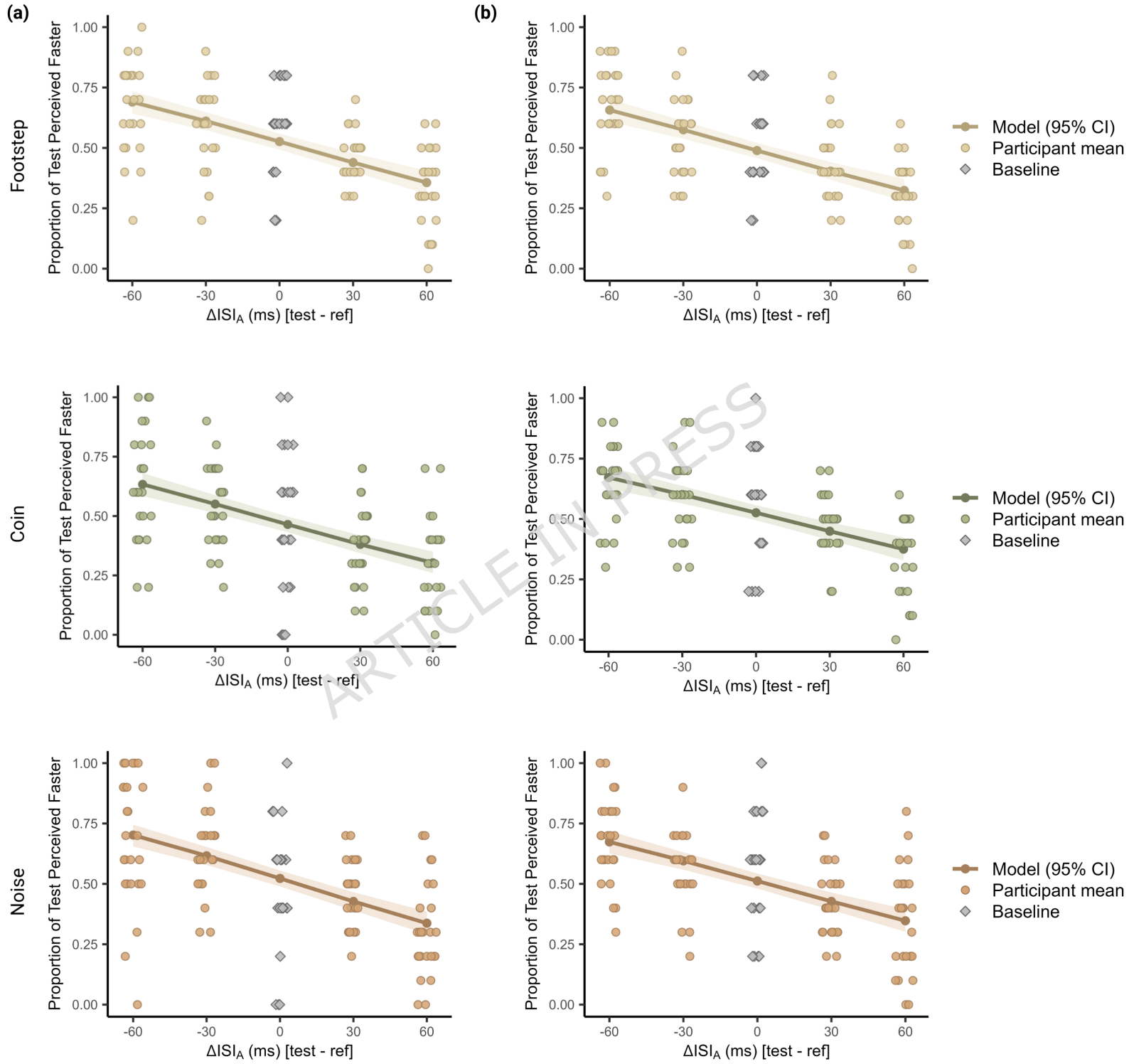
ARTICLE IN PRESS

ARTICLE IN PRESS





ARTICLE IN PRESS



ARTICLE IN PRESS