Title page

Cortical tracking of hierarchical rhythms orchestrates the multisensory processing of biological motion

Li Shen[1, 2], Shuo Li[1, 2], Yuhao Tian[1, 2], Ying Wang[1, 2*], Yi Jiang[1, 2]

[1] State Key Laboratory of Brain and Cognitive Sciences, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

[2] Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

***Corresponding author**: Ying Wang (wangying@psych.ac.cn)

Short title: Cortical tracking of audiovisual biological motion

Abstract

When observing others' behaviors, we continuously integrate their movements with the corresponding sounds to enhance perception and develop adaptive responses. However, how the human brain integrates these complex audiovisual cues based on their natural temporal correspondence remains unknown. Using electroencephalogram, we demonstrated that rhythmic cortical activity tracked the hierarchical rhythmic structures in audiovisually congruent human walking movements and footstep sounds. Remarkably, the cortical tracking effects at different time scales exhibit distinct modes of multisensory integration: an additive mode in a basic-level, narrower temporal integration window (step-cycle) and a super-additive enhancement in a higher-order, broader temporal window (gait-cycle). Moreover, only the cortical tracking of higher-order rhythmic structures is specialized for the multisensory integration of human motion signals and correlates with individuals' autistic traits, suggesting its functional relevance to biological motion perception and social cognition. These findings unveil the multifaceted roles of entrained cortical activity in the multisensory perception of human motion, shedding light on how neural coding of hierarchical temporal structures orchestrates the processing of complex, rhythmic stimuli in natural contexts.

Keywords: multisensory integration, biological motion, rhythm, cortical tracking, neural entrainment

Introduction

The perception of biological motion (BM), the movements of living creatures, is a fundamental ability of the human visual system. Extensive evidence shows that humans can readily perceive BM from a visual display depicting just a handful of light dots attached to the head and major joints of a moving person (Blake & Shiffrar, 2007). Nevertheless, in real life, BM perception often occurs in multisensory contexts. For instance, one may simultaneously hear footstep sounds while seeing others walking. The integration of these visual and auditory BM cues facilitates the detection, discrimination, and attentional processing of BM (Mendonça et al., 2011; Shen, Lu, Wang, et al., 2023; Thomas & Shiffrar, 2013; van der Zwan et al., 2009). Notably, such benefits are diminished when the visual BM is deprived of characteristic kinematic cues but not low-level motion attributes (Brooks et al., 2007; Shen, Lu, Wang, et al., 2023; Thomas & Shiffrar, 2010), and the temporal windows of perceptual audiovisual synchrony are different between BM and non-BM stimuli (Arrighi et al., 2006; Saygin et al., 2008), highlighting the specificity of audiovisual BM processing. This specificity may relate to the evolutionary significance of BM and its relevance in social situations. In particular, integrating multisensory BM cues is foundational for perceiving and attending to other people and developing further social interaction. Such ability is usually compromised in people with social deficits, such as individuals with autism spectrum disorder (ASD) (Feldman et al., 2018), and even in non-clinical populations with high autistic traits (Ujiie et al., 2015). These findings underline the unique contribution of multisensory BM processing to human

perception and social cognition. However, despite the behavioral evidence, the neural encoding of audiovisual BM cues and its possible link with individuals' social cognitive capability remains largely unexplored.

An intrinsic property of human movements (such as walking and running) is that they are rhythmic and accompanied by frequency-congruent sounds. The audiovisual integration (AVI) of such rhythmic stimuli may involve cortical entrainment, a process that brain activity aligning with and tracking external rhythms, revealed by increased power or phase coherence of neural oscillations at corresponding frequencies (Bauer et al., 2020; Ding et al., 2016; Obleser & Kayser, 2019). Studies based on simple or discrete stimuli show that temporal congruency in auditory and visual rhythms significantly enhances the cortical tracking of rhythmic stimulations in both modalities (Covic et al., 2017; Keitel & Müller, 2016; Nozaradan et al., 2012b). Unlike these stimuli, BM conveys complex hierarchical rhythmic structures corresponding to integration windows at multiple temporal scales. For example, the human locomotion movement has a narrower integration window consisting of each step (i.e., step cycle) and a broader integration window incorporating the opponent motion of the two feet (i.e., gait cycle). A recent study suggests that neural tracking of these nested kinematic structures contributes to the spatiotemporal integration of visual BM cues in different manners (Shen, Lu, Yuan, et al., 2023). However, it remains open whether and how the cortical tracking of hierarchical rhythmic structures underpins the AVI of BM information.

To tackle this issue, we recorded electroencephalogram (EEG) signals from participants who viewed rhythmic point-light walkers or/and listened to the corresponding footstep sounds under visual (V), auditory (A), and audiovisual (AV) conditions in Experiments 1a & 1b (**Fig. 1**). An enhanced cortical tracking effect in the AV condition compared to each unisensory condition will indicate significant multisensory gains. Moreover, we adopted an additive model to classify multisensory integration based on the AV vs A+V comparison. This model assumes independence between inputs from each sensory modality and distinguishes among sub-additive (AV < A+V), additive (AV = A+V), and super-additive (AV > A+V) response modes (see a review by Stevenson et al., 2014). The additive mode represents a linear combination between two modalities. In contrast, the super-additive and sub-additive modes indicate non-linear interaction processing, either with potentiated neural activation to facilitate the perception or detection of near-threshold signals (super-additive) or a deactivation mechanism to minimize the processing of redundant information cross-modally (sub-additive) (Laurienti et al., 2005; Metzger et al., 2020; Stanford et al., 2005; Wright et al., 2003).

Experiment 2 examined to what extent the AVI effect was specific to the multisensory processing of BM by using non-BM (inverted visual stimuli) as a control. Inversion disrupts the unique, gravity-compatible kinematic features of BM but not the rhythmic signals generated by low-level motion cues (Ma et al., 2022; Shen, Lu, Yuan, et al., 2023; Simion et al., 2008; Troje & Westhoff, 2006; Wang et al., 2022), thus is expected to interfere with the BM-specific neural processing. Participants perceived

visual BM stimuli accompanied by temporally congruent or incongruent BM sounds. Comparing the congruency effect in neural responses between the upright and inverted conditions allowed us to verify whether the AVI of BM involves a mechanism distinct from that underlies the AVI of non-BM. In addition, to further explore the functional relevance of the BM-specific neural tracking effect, we examined its potential relationship with observers' autistic traits. Previous behavioral studies found reduced orienting to audiovisually synchronized BM stimuli in ASD (Falck-Ytter et al., 2018; Klin et al., 2009). Since social cognitive deficits in ASD lie on a continuum extending from the clinical to nonclinical populations with different levels of autistic traits, as measured by the Autism-Spectrum Quotient (AQ) (Baron-Cohen et al., 2001), here we investigated the correlation between the cortical tracking of audiovisual BM and the individuals' AQ scores.

Results

In all experiments, 17%–23% of the trials were randomly selected as catch trials, in which the color of the walker changed one or two times throughout the trial, and there was no color change in other trials. Participants were required to detect the color change of visual stimuli (0-2 times during one trial) to maintain attention. Behavioral analysis on all trials showed that their performances for the task were generally high and equally well in all conditions of Experiment 1a (mean accuracy > 98%; F (2, 46) = 0.814, p = .450, η_p^2 = 0.034), Experiment 1b (mean accuracy > 98%; F (2, 46) = 0.615, p = .545, η_p^2 = 0.026), and Experiment 2 (mean accuracy > 98%; F (3, 69) =

0.493, p = .688, η_p^2 = 0.021), indicating comparable attention state across conditions. The catch trials were excluded from the following EEG analysis.

Cortical tracking of rhythmic structures in audiovisual BM reveals AVI

Experiment 1a. In Experiment 1a, we examined the cortical tracking of rhythmic BM information under V, A, and AV conditions (**Fig. 1c**). We were interested in two critical rhythmic structures in the walking motion sequence, i.e., the gait cycle and the step cycle (**Fig. 1a & 1b**). During walking, each step of the left or right foot occurs alternatively to form a step cycle, and the antiphase oscillations of limbs during two steps characterize a gait cycle (Shen, Lu, Yuan, et al., 2023). In Experiment 1a, the frequency of a full gait cycle is 1 Hz, and the step-cycle frequency is 2 Hz. The strength of the cortical tracking effect was quantified by the amplitude peaks emerging from the EEG spectra at these frequencies.

As shown in the grand average amplitude spectra (**Fig. 2a**), both the responses in three conditions showed clear peaks at step-cycle frequency (2 Hz; V: t (23) = 6.963, p < 0.001; A: t (23) = 6.073, p < .001; AV: t (23) = 7.054, p < 0.001; FDR corrected). In contrast, at gait-cycle frequency (1 Hz), only the response to AV stimulation showed significant peaks (V: t (23) = -2.072, p = 0.975; A: t (23) = -0.054, p = 0.521; AV: t (23) = 4.059, p < 0.001; FDR corrected). Besides, we also observed significant peaks at 4 Hz in all three conditions (ps < 0.001, FDR corrected), which showed a similar audiovisual integration mode as 2 Hz (see more details in Supplementary Information).

Furthermore, we directly compared the cortical tracking effects between different conditions via a two-tailed paired t-test. At both 1 Hz (**Fig. 2b**) and 2 Hz (**Fig. 2c**), the amplitude in the AV condition was greater than that in the V condition (1 Hz: t (23) = 4.664, p < 0.001, Cohen's d = 0.952; 2 Hz: t (23) = 5.132, p < 0.001, Cohen's d = 1.048) and the A condition (1 Hz: t (23) = 2.391, p = 0.025, Cohen's d = 0.488; 2 Hz: t (23) = 3.808, p < 0.001, Cohen's d = 0.777), respectively, suggesting multisensory gains. More importantly, at 1Hz, the amplitude in the AV condition was significantly larger than the algebraic sum of those in the A and V conditions (t (23) = 3.028, p = 0.006, Cohen's d = 0.618), indicating a super-additive audiovisual integration effect. While at 2Hz, the amplitude in the AV condition was comparable to the unisensory sum (t (23) = -0.623, p = 0.539, Cohen's d = -0.127), indicating additive audiovisual integration.

Experiment 1b. To further test whether such cortical tracking effect can apply to stimuli with a different speed, Experiment 1b altered the frequencies of the gait cycle and the corresponding step cycle to 0.83 Hz and 1.67 Hz while adopting the same paradigm as Experiment 1a. Consistent with Experiment 1a, the frequency-domain analysis revealed significant cortical tracking of the audiovisual stimuli at the new speeds. As shown in **Fig. 2d**, both the responses to V, A, and AV stimuli showed clear peaks at step-cycle frequency (1.67 Hz; V: t (23) = 3.473, p = .001; A: t (23) = 9.194, p < .001; AV: t (23) = 8.756, p < .001; FDR corrected) and its harmonics (3.33 Hz, ps < .001, FDR corrected). In contrast, at gait-cycle frequency (0.83 Hz), only the response to AV stimuli showed significant peaks (V: t (23) = -1.125, p = .846; A: t (23)

= -2.449, p = .989; AV: t (23) = 3.052, p = .003; FDR corrected).

At both 0.83 Hz (**Fig. 2e**) and 1.67 Hz (**Fig. 2f**), the amplitude in the AV condition was stronger or marginally stronger than that in the V condition (0.83 Hz: t (23) = 2.665, p = .014, Cohen's d = 0.544; 1.67 Hz: t (23) = 6.380, p < .001, Cohen's d = 1.302) and the A condition (0.83 Hz: t (23) = 3.625, p < .001, Cohen's d = 0.740; 1.67 Hz: t (23) = 1.752, p = .093, Cohen's d = 0.358), respectively, suggesting multisensory gains. More importantly, at 0.83 Hz, the amplitude in the AV condition was significantly larger than the sum of those in the A and V conditions (t (23) = 3.240, p = .004, Cohen's d = 0.661), indicating a super-additive audiovisual integration effect. While at 1.67 Hz, the amplitude in the AV condition was comparable to the unisensory sum (t (23) = -0.735, p = .470, Cohen's d = -0.150), indicating linear audiovisual integration. Significant peaks were also observed at 3.33 Hz in all three conditions (ps < 0.001, FDR corrected), which showed similar audiovisual integration mode as 1.67 Hz (see more details in Supplementary Information).

In summary, results from Experiments 1a & 1b consistently showed that the cortical tracking of the audiovisual signals at different temporal scales exhibit distinct audiovisual integration modes, i.e., the super-additive effect at gait-cycle frequency and the additive effect at step-cycle frequency, indicating that the cortical tracking effects at the two temporal scales might be driven by functionally dissociable mechanisms.

Cortical tracking of higher-order rhythmic structure contributes to the AVI of BM

To further explore whether and how the cortical tracking of rhythmic structures contributes to the specialized audiovisual processing of BM, both upright and inverted BM stimuli were adopted in Experiment 2. The task and the frequencies of visual stimuli in Experiment 2 were same as Experiment 1a. Specifically, participants were required to perform the change detection task when perceiving upright and inverted visual BM sequences (1 Hz for gait-cycle frequency and 2 Hz for step-cycle frequency) accompanied by frequency congruent (1 Hz) or incongruent (0.6 Hz and 1.4 Hz) footstep sounds (**Fig. 1c**). The audiovisual congruency effect, characterized by stronger neural responses in the audiovisual congruent condition compared with the incongruent condition, can be taken as an index of AVI (Fleming et al., 2020; Jones & Jarick, 2006; Maddox et al., 2015; Wuerger, Crocker-Buque, et al., 2012). A stronger congruency effect in the upright condition relative to the inverted condition characterizes an AVI process specific to BM information.

We calculated the audiovisual congruency effect for the upright and the inverted conditions, respectively. Then, we contrasted the congruency effect between the upright and inverted conditions to search for clusters showing a significant difference, which equaled identifying an interaction effect, using a cluster-based permutation test over all electrodes (n = 1000, *alpha* = 0.05; see *Methods*). At 1 Hz, the congruency effect in the upright condition was significantly stronger than that in the inverted

condition in a cluster at the right hemisphere (**Fig. 3a**, lower panel, p = 0.029; C2, CPz, CP2, CP4, CP6, Pz, P2, P4, P6), revealing a BM-specific AVI process. Then we averaged the amplitude of electrodes within the significant cluster and further performed two-tailed paired t-tests to examine whether the congruency was significant in the upright and the inverted conditions, respectively. Results showed that (**Fig. 3b**) audiovisually congruent BM information enhanced the oscillatory amplitude relative to the incongruent ones only for upright BM stimuli (t (23) = 4.632, p < 0.001, Cohen's d = 0.945) but not when visual BM was inverted (t (23) = 0.480, p = 0.635, Cohen's d = 0.098).

In contrast, at 2 Hz, no cluster showed a significantly different congruency effect between the upright and inverted conditions (**Fig. 3d**), suggesting no BM-specific AVI process. We then conducted further analysis based on the electrodes yielded by 1 Hz as marked in **Fig. 3a**. Results showed that both upright and inverted stimuli induced a significant congruency effect at 2 Hz (**Fig. 3e**; Upright: t (23) = 3.096, p = 0.005, Cohen's d = 0.632; Inverted: t (23) = 2.672, p = 0.014, Cohen's d = 0.545).

To verify the dissociation of the BM-specific AVI effects between the gait-cycle frequency and the step-cycle frequency, we performed a three-way repeated-measures ANOVA with cycle frequency (1 Hz vs. 2 Hz), orientation (upright vs. inverted), and audiovisual congruency (congruent vs. incongruent) as within-subject factors. We converted the amplitude data at each frequency condition into Z-scores to reduce the potential influence of amplitude magnitude difference between the two

frequencies. The analysis of these data revealed a significant three-way interaction (F (1,23) = 7.501, p = 0.012, η_p^2 = 0.246), further supporting that the audiovisual integration processing of BM is different between 1 Hz and 2 Hz.

BM-specific cortical tracking correlates with autistic traits

Furthermore, we examined the link between individuals' autistic traits and the neural responses underpinning the AVI of BM, measured by the difference of congruency effect between the upright and the inverted BM conditions, using Pearson correlation analysis. After removing one outlier (whose neural response exceeded 3 SD from the group mean), we observed an evident negative correlation between individuals' AQ scores and their neural responses at 1 Hz (**Fig. 3c**, r = -0.493, p = 0.017) but not at 2 Hz (**Fig. 3f**, r = -0.158, p = .460). The lack of significant results at 2 Hz was not attributable to electrode selection bias based on the significant cluster at 1 Hz, as similar results were observed when we performed the analysis on clusters showing non-selective significant congruency effects at 2 Hz (see the control analysis in Supplementary Information for details). Besides, we split the participants based on their median AQ score and found that, compared with the high AQ group, the low AQ group showed a greater BM-specific cortical tracking effect at 1 Hz but not at 2 Hz. These findings provide further support to the functional relevance between social cognition and cortical tracking of biological motion as well as its dissociation at the two temporal scales (see more details in Supplementary Information).

Discussion

The current study investigated the neural implementation for the AVI of human BM information and its functional implications. We found that, even under a motion-irrelevant color detection task, observers' neural activity tracked the temporally corresponding audiovisual BM signals at the frequencies of two rhythmic structures, i.e., the higher-order structure of gait cycle at a larger integration window and the basic-level structure of step cycle at a smaller integration window. The strength of these cortical tracking effects was enhanced under the audiovisual condition than in the visual-only or auditory-only condition, indicating multisensory gains. More crucially, although the cortical tracking of both gait-cycle and step-cycle gain benefits from multisensory correspondence, the mechanisms underlying these two processes appear to be different. At step-cycle frequency, the cortical tracking effect in the AV condition equaled the additive sum of the unisensory conditions. Such linear integration may result from concurrent, independent processing of unisensory inputs without additional interaction of them (Stein et al., 2009). In contrast, at gait-cycle frequency, the congruent audiovisual signals led to a super-additive multisensory enhancement over the linear combination of auditory and visual conditions (AV > A+V), despite that there was no evident cortical tracking effect in the visual condition, different from previous findings obtained with a motion-relevant change detection task (Shen, Lu, Yuan, et al., 2023). This super-additive multisensory enhancement may bring about decreased thresholds of detection and identification (Stanford et al.,

2005), allowing us to achieve a more clear and stable perception of the external environment and detect weak stimulus changes in time and respond adaptively.

Furthermore, results from Experiment 2 demonstrated that the cortical tracking of rhythmic structure corresponding to the gait cycle rather than the step cycle is relevant to the specialized processing of audiovisual BM information. In particular, the AVI effect at step-cycle frequency was significant for both upright and inverted BM signals and comparable between the two conditions, while the AVI effect at gait-cycle frequency was only significant in the upright condition and was greater than that in the inverted condition. The inversion effect has long been regarded as a marker of the specificity of BM processing in numerous behavioral and neuroimaging studies (Grossman & Blake, 2001; Ma et al., 2022; Shen, Lu, Yuan, et al., 2023; Simion et al., 2008; Troje & Westhoff, 2006; Vallortigara & Regolin, 2006; Wang et al., 2014; Wang & Jiang, 2012; Wang et al., 2022). Our current findings of the inversion effect in the cortical tracking of audiovisual BM at the gait-cycle suggest that the neural encoding of the higher-order rhythmic structure reflects the AVI of BM and contributes to the specialized processing of BM information. In contrast, the cortical tracking of the step-cycle may reflect the integration of basic motion signals and corresponding sounds. Together, these results reveal that the neural tracking of different levels of kinematic structures plays distinct roles in the AVI of BM, which may result from the interplay of stimulus-driven and domain-specific mechanisms. In addition, a recent study demonstrated that listening to frequency-congruent footstep sounds, compared with incongruent sounds, enhanced the visual search for human walkers but not for

non-BM stimuli containing the same rhythmic signals, indicating that audiovisual correspondence specifically enhances the perceptual and attentional processing of BM (Shen, Lu, Wang, et al., 2023). Future research could examine whether the cortical tracking of rhythmic structures plays a functional role in this process, which may shed more light on the behavioral relevance of the cortical tracking effect to BM perception.

Besides the temporal dynamics of neural activity revealed by the cortical tracking process, we found that the BM-specific AVI effect was associated with neural activity in the right temporoparietal electrodes. This finding likely relates to the activation of the right posterior superior temporal sulcus (pSTS), a region responding to both auditory and visual BM information and being causally involved in BM perception (Bidet-Caulet et al., 2005; Grossman et al., 2005; Wang et al., 2022). While previous fMRI studies have observed STS activation when processing spatial or semantic correspondence between audiovisual BM (Meyer et al., 2011; Wuerger, Parkes, et al., 2012), whether this region also engages in the audiovisual processing of BM signals based on temporal correspondence remains unknown. The current study provides preliminary evidence for such a possibility, inviting future research to localize the exact source of the multisensory integration processes based on imaging data with high spatial and temporal resolutions, such as MEG.

Cortical tracking of external rhythms is also described as cortical entrainment in a broad sense (Ding et al., 2016; Obleser & Kayser, 2019). Controversy remains

regarding the involvements of endogenous neural oscillations and stimulus-evoked responses in these processes (Duecker et al., 2024), as it is challenging to fully dissociate these components due to their intricate interplay (Herrmann et al., 2016; Hosseinian et al., 2021). Previous research has demonstrated that cortical tracking or entrainment plays a role in the multisensory processing of simple or discrete rhythmic signals (Covic et al., 2017; Keitel & Müller, 2016; Nozaradan et al., 2012b). These findings may partially explain the non-selective AVI effect at the step-cycle in the current study. However, we found that the cortical tracking of the higher-order rhythmic structure formed by spatiotemporal integration of meaningful BM information (i.e., the gait cycle of upright walkers rather than inverted walkers) is selectively engaged by the AVI of BM, suggesting that the multisensory processing of natural continuous stimuli may involve unique mechanisms besides the purely stimulus-driven AVI process. These findings provide significant implications for the neural processing of natural audiovisual stimuli with complex temporal structures. Similar to BM, other natural rhythmic stimuli, like speech and music, also convey hierarchical structures that can entrain neural oscillations at different temporal scales, both in unisensory (Ding et al., 2016; Doelling & Poeppel, 2015) and multisensory contexts (Biau et al., 2022; Crosse et al., 2015; Nozaradan et al., 2016). Possibly, the audiovisual processing of these stimuli also engages multi-scale neural coding mechanisms that play distinct functions in perception. Investigating this issue and comparing the results with BM studies will help complete the picture of how the human brain integrates complex, rhythmic information sampled from different

sensory modalities to orchestrate perception in a natural scenario.

Last but not least, our study demonstrated that the selective cortical tracking of higher-level rhythmic structure in audiovisually congruent BM signals negatively correlated with individual autistic traits. This finding highlights the functional significance of the neural tracking of audiovisual BM signals in social cognition. It also offers the first evidence that differences in audiovisual BM processing are already present in nonclinical individuals and associated with their autistic traits, beyond previous evidence for atypical audiovisual BM processing in ASD populations (Falck-Ytter et al., 2013, 2018; Klin et al., 2009), lending support to the continuum view of ASD (Baron-Cohen et al., 2001). Meanwhile, given that impaired audiovisual BM processing at the early stage may influence social development and result in cascading consequences for lifetime impairments in social interaction (Falck-Ytter et al., 2018; Klin et al., 2005), it is worth exploring neural tracking of audiovisual BM signals in children with different autistic levels, which may help reveal whether deficits in such ability could serve as an early neural hallmark for ASD.

Materials and Methods

Participants

Seventy-two participants (mean age ± SD = 22.4 ± 2.6 years, 35 females) took part in the study, 24 for each of Experiment 1a, Experiment 1b, and Experiment 2. All of them had normal or corrected-to-normal vision and reported no history of neurological, psychiatric, or hearing disorders. They were naïve to the purpose of the study and gave informed consent according to procedures and protocols approved by the institutional review board of the Institute of Psychology, Chinese Academy of Sciences.

Stimuli

Visual stimuli. The visual stimuli (**Fig. 1a**, left panel) consisted of 13 point-light dots attached to the head and major joints of a human walker (Vanrie & Verfaillie, 2004). The point-light walker was presented at the center of the screen without translational motion. It conveys rhythmic structures specified by recurrent forward motions of bilateral limbs (**Fig. 1a**, right panel). Each step, regardless of left or right foot, occurs recurrently to form a step cycle. The antiphase oscillations of limbs during two steps characterize a gait cycle (Shen, Lu, Yuan, et al., 2023). In Experiment 1a, a full gait cycle took 1 second and was repeated 6 times to form a 6-second walking sequence. That is, the gait-cycle frequency is 1 Hz and the step-cycle frequency is 2 Hz. In Experiment 1b, the gait-cycle frequency was 0.83 Hz and the step-cycle frequency

was 1.67 Hz. The gait cycle was repeated 6 times to form a 7.2-second walking sequence. The stimuli in Experiment 2 were the same as that in Experiment 1a. Meanwhile, the point-light BM was mirror-flipped vertically to generate inverted BM (**Fig. 1a**, left panel), which preserves the temporal structure of the stimuli but distorts its distinctive kinematic features, such as movement that is compatible with the effect of gravity (Shen, Lu, Yuan, et al., 2023; Troje & Westhoff, 2006; Wang et al., 2022).

Auditory stimuli. Auditory stimuli were continuous footstep sounds (6 s) with a sampling rate of 44,100 Hz. As shown in **Fig. 1b**, in Experiments 1a & 2, the gait-cycle frequency of congruent sounds was 1 Hz, which consisted of two steps or two impulses generated by each foot striking the ground within one gait cycle. The incongruent sounds included a faster (1.4 Hz) and a slower (0.60 Hz) sound. Both congruent and incongruent sounds were generated by manipulating the temporal interval between two successive impulses based on the same auditory stimuli. In Experiment 1b, the gait-cycle frequency of sound was 0.83 Hz.

Stimuli presentation. The visual stimuli were rendered white against a grey background and displayed on a CRT (cathode ray tube) monitor. Participants sat 60 cm from the computer screen (1280×1024 at 60 Hz; High: 37.5 cm; Width: 30 cm), with their heads held stationary on a chinrest. The auditory stimuli were presented binaurally over insert earphones. All stimuli were generated and presented using MATLAB together with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Procedure and task

Experiment 1a. The experiment was conducted in an acoustically dampened and electromagnetically shielded chamber. Participants completed the task under three conditions (Visual: V; Auditory: A; Audiovisual: AV) with the same procedure (**Fig. 1c**) except for the stimuli. In the V condition, each trial began with a white fixation cross (0.42° × 0.42°) displayed at the center of a gray background for a random duration (0.8 s to 1 s). Subsequently, a 6-s point-light walker (3.05°×5.47°) walked toward the left or right at a constant walking cycle frequency (1 Hz). To maintain observers' attention, 17%–23% of the trials were randomly selected as catch trials, in which the color of the walker changed (the RGB values changed from [255 255 255] to [207 207 207]) one or two times throughout the trial. Each change lasted 0.5 s. Observers were required to report the number of changes (0, 1, or 2) via keypresses as accurately as possible after the point-light display was replaced by a red fixation. The next trial started 2–3 s after the response. In the A condition, the 6 s-stimuli were replaced by a visually static BM figure accompanied by continuous footstep sounds. The frequency of footstep sounds was congruent with the frequency of visual BM in the V condition. In the AV condition, the stimuli were temporally congruent visual BM sequences (as in the V condition) and footstep sounds (as in the A condition). Three conditions were conducted in separate blocks. V condition was performed in the middle of A and AV conditions. The order of A and AV conditions was counterbalanced across participants. Each participant completed 40 experimental trials without changes and 10-15 catch trials in each condition, resulting in a total of

150-165 trials. In each condition, participants completed a practice session with 3 trials to get familiar with the task before the formal EEG experiment.

Experiment 1b. The procedure of Experiment 1b was the same as that for Experiment 1a but with two exceptions. First, to test if the cortical tracking effect can apply to stimuli with a different speed, we altered the frequencies of gait and step cycles to 0.83 Hz and 1.67 Hz. Second, we presented the 3 conditions (V, A, and AV) in a completely random order to eliminate the influence of presentation order. To minimize the potential influence of condition switch, we increased the trial number in the practice session from 3 to 14 for each condition.

Experiment 2. The procedure in Experiment 2 was similar to the AV condition in Experiment 1a, except that the visually displayed BM was accompanied by frequency congruent (1 Hz) or incongruent (0.6 or 1.4 Hz) footstep sounds. Each participant completed a total of 76 experiment trials, consisting of 36 congruent-trials, 20 incongruent-trials with faster sounds (1.4 Hz), and 20 incongruent-trials with slower sounds (0.6 Hz). These trials were assigned to 3 blocks based on the frequency of the footstep sounds, with the order of the three frequencies balanced across participants. Besides, an inverted BM was used as a control to investigate whether there is a specialized mechanism tuned to the AVI of life motion signals. The order of upright and inverted conditions was balanced across participants. Meanwhile, we measured the participants' autistic traits by using the Autism-Spectrum Quotient, or AQ questionnaire (Baron-Cohen et al., 2001). Higher AQ scores indicate a higher

level of autistic traits.

EEG recording and analysis

EEG was recorded at 1000 Hz using a SynAmps^2 NeuroScan amplifier System with 64 electrodes placed on the scalp according to the international 10-20 system. Horizontal and vertical eye movements were measured via four additional electrodes placed on the outer canthus of each eye and the inferior and superior areas of the left orbit. Impedances were kept below 5 kΩ for all electrodes.

Preprocessing. The catch trials were excluded from EEG analysis. All preprocessing and further analyses were performed using the FieldTrip toolbox (Oostenveld et al., 2011; http://fieldtriptoolbox.org) in the MATLAB environment. EEG recordings were pass-filtered between 0.1 and 30 Hz, and down-sampled to 100 Hz. Then the continuous EEG data were cut into epochs ranging from -1s to 6 gait cycles (7.2 s in Experiment 1b and 6 s in other experiments) time-locked to the onset of the visual point-light stimuli. The epochs were visually inspected, and trials contaminated with excessive noise were excluded from the analysis. After the trial rejection, eye and cardiac artifacts were removed via independent component analysis based on the Runica algorithm (Bell & Sejnowski, 1995; Jung et al., 2000; Makeig, 2002). Then the cleaned data were re-referenced to the average mastoids (M1 and M2). To minimize the influence of stimulus-onset evoked activity on EEG spectral decomposition, the EEG recording before the onset of the stimulus and the first cycle (1 s in Experiments 1a & 2; 1.2 s in Experiment 1b) of each trial was excluded (Nozaradan et al., 2012a).

After that, the EEG epochs were averaged across trials for each participant and condition.

Frequency-Domain analysis and statistics. A Fast Fourier Transform (FFT) with zero padding (1200) was used to convert the averaged EEG signals from the temporal domain to the spectral domain, resulting in a frequency resolution of 0.083 Hz, i.e., 1/12 Hz, which is sufficient for observing neural responses around the frequency of the rhythmic BM structures in all experiments. When performing FFT, a Hanning window was adopted to minimize spectral leakage. Then, to remove the 1/f trend of the response amplitude spectrum and identify spectral peaks, the response amplitude at each frequency was normalized by subtracting the average amplitude measured at the neighboring frequency bins (two bins on each side) (Nozaradan et al., 2012a). We calculated the normalized amplitude separately for each electrode (except for electrooculogram electrodes, CB1, and CB2), participant, and condition.

In Experiment 1, the normalized amplitude in all electrodes was averaged and a right-tailed one-sample t-test against zero was performed on the grand average amplitude to test whether the neural response in each frequency bin showed a significant tracking effect or spectral peak. This test was applied to all frequency bins below 5.33 Hz and multiple comparisons were controlled by false discovery rate (FDR) correction at $p < 0.05$ (Benjamini & Hochberg, 1995). In Experiment 2, to further identify the BM-specific AVI process, the audiovisual congruency effect was compared between the upright and inverted conditions using a cluster-based

permutation test over all electrodes (1000 iterations, requiring a cluster size of at least 2 significant neighbors, a two-sided t-test at $p < 0.05$ on the clustered data) (Oostenveld et al., 2011; http://fieldtriptoolbox.org). This allowed us to identify the spatial distribution of the BM-specific congruency effect.

Acknowledgments

This research was supported by grants from the Ministry of Science and Technology of China (STI2030-Major Projects 2021ZD0204200 and 2021ZD0203800), the National Natural Science Foundation of China (32171059 and 31830037), the Interdisciplinary Innovation Team (JCTD-2021-06), the Youth Innovation Promotion Association of the Chinese Academy of Sciences, the Fundamental Research Funds for the Central Universities, and the China Postdoctoral Science Foundation (2024M170993 and 2024M753476).

Authors Contributions

Li Shen: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing-original draft, Writing–review & editing. **Shuo Li & Yuhao Tian**: Investigation, Writing-original draft. **Ying Wang**: Conceptualization, Methodology, Supervision, Writing–review & editing. **Yi Jiang**: Conceptualization, Supervision, Writing–review & editing.

Conflict of interest declaration

The authors declare no conflicts of interest.

Data availability

The supplementary information files, data, and code accompanying this study are made available at https://www.scidb.cn/en/s/jeiiea.

References

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/a:1005653411471

Bauer, A.-K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of Neural Oscillations and Cross-modal Influences. *Trends in Cognitive Sciences*, *24*(6), 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Bell, A. J., & Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Biau, E., Schultz, B. G., Gunter, T. C., & Kotz, S. A. (2022). Left Motor δ Oscillations Reflect Asynchrony Detection in Multisensory Speech Perception. *Journal of Neuroscience*, *42*(11), 2313–2326. https://doi.org/10.1523/JNEUROSCI.2965-20.2022

Bidet-Caulet, A., Voisin, J., Bertrand, O., & Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *NeuroImage*, *28*(1), 132–139.

https://doi.org/10.1016/j.neuroimage.2005.06.018

Blake, R., & Shiffrar, M. (2007). Perception of Human Motion. *Annual Review of Psychology*, *58*(1), 47–73. https://doi.org/10.1146/annurev.psych.57.102904.190152

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

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

Covic, A., Keitel, C., Porcu, E., Schröger, E., & Müller, M. M. (2017). Audio-visual synchrony and spatial attention enhance processing of dynamic visual stimulation independently and in parallel: A frequency-tagging study. *NeuroImage*, *161*, 32–42. https://doi.org/10.1016/j.neuroimage.2017.08.022

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *Journal of Neuroscience*, *35*(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, *112*(45), E6233–E6242. https://doi.org/10.1073/pnas.1508431112

Duecker, K., Doelling, K. B., Breska, A., Coffey, E. B. J., Sivarao, D. V., & Zoefel, B. (2024). Challenges and approaches in the study of neural entrainment. *Journal of Neuroscience*,

44(40). https://doi.org/10.1523/JNEUROSCI.1234-24.2024

Falck-Ytter, T., Nyström, P., Gredebäck, G., Gliga, T., Bölte, S., & the EASE team. (2018). Reduced orienting to audiovisual synchrony in infancy predicts autism diagnosis at 3 years of age. *Journal of Child Psychology and Psychiatry*, *59*(8), 872–880. https://doi.org/10.1111/jcpp.12863

Falck-Ytter, T., Rehnberg, E., & Bölte, S. (2013). Lack of Visual Orienting to Biological Motion and Audiovisual Synchrony in 3-Year-Olds with Autism. *PLoS ONE*, *8*(7), e68816. https://doi.org/10.1371/journal.pone.0068816

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Fleming, J. T., Noyce, A. L., & Shinn-Cunningham, B. G. (2020). Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. *Neuropsychologia*, *146*, 107530. https://doi.org/10.1016/j.neuropsychologia.2020.107530

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, *45*(22), 2847–2853. https://doi.org/10.1016/j.visres.2005.05.027

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, *41*(10), 1475–1482. https://doi.org/10.1016/S0042-6989(00)00317-5

Herrmann, C. S., Murray, M. M., Ionta, S., Hutt, A., & Lefebvre, J. (2016). Shaping Intrinsic

Neural Oscillations with Periodic Stimulation. *Journal of Neuroscience*, *36*(19), 5328–5337. https://doi.org/10.1523/JNEUROSCI.0236-16.2016

Hosseinian, T., Yavari, F., Biagi, M. C., Kuo, M.-F., Ruffini, G., Nitsche, M. A., & Jamil, A. (2021). External induction and stabilization of brain oscillations in the human. *Brain Stimulation*, *14*(3), 579–587. https://doi.org/10.1016/j.brs.2021.03.011

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, *174*(3), 588–594. https://doi.org/10.1007/s00221-006-0634-0

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745–1758. https://doi.org/10.1016/S1388-2457(00)00386-2

Keitel, C., & Müller, M. M. (2016). Audio-visual synchrony and feature-selective attention co-amplify early visual processing. *Experimental Brain Research*, *234*(5), 1221–1231. https://doi.org/10.1007/s00221-015-4392-8

Klin, A., Jones, W., Schultz, R. T., & Volkmar, F. R. (2005). The Enactive Mind-From Actions to Cognition: Lessons from Autism. In *Handbook of autism and pervasive developmental disorders: Diagnosis, development, neurobiology, and behavior, Vol. 1, 3rd ed* (pp. 682–703). John Wiley & Sons Inc.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, *459*(7244), 257–261. https://doi.org/10.1038/nature07868

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, *166*(3), 289–297. https://doi.org/10.1007/s00221-005-2370-2

Ma, X., Yuan, X., Liu, J., Shen, L., Yu, Y., Zhou, W., Liu, Z., & Jiang, Y. (2022). Gravity-Dependent Animacy Perception in Zebrafish. *Research*, *2022*. https://doi.org/10.34133/2022/9829016

Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife*, *4*, e04995. https://doi.org/10.7554/eLife.04995

Makeig, S. (2002). Response: Event-related brain dynamics – unifying brain electrophysiology. *Trends in Neurosciences*, *25*(8), 390. https://doi.org/10.1016/S0166-2236(02)02198-7

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

Metzger, B. A., Magnotti, J. F., Wang, Z., Nesbitt, E., Karas, P. J., Yoshor, D., & Beauchamp, M. S. (2020). Responses to Visual Speech in Human Posterior Superior Temporal Gyrus Examined with iEEG Deconvolution. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *40*(36), 6938–6948. https://doi.org/10.1523/JNEUROSCI.0279-20.2020

Meyer, G. F., Greenlee, M., & Wuerger, S. (2011). Interactions between auditory and visual semantic stimulus classes: Evidence for common processing networks for speech and

body actions. *Journal of Cognitive Neuroscience*, *23*(9), 2291–2308. https://doi.org/10.1162/jocn.2010.21593

Nozaradan, S., Peretz, I., & Mouraux, A. (2012a). Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm. *Journal of Neuroscience*, *32*(49), 17572–17581. https://doi.org/10.1523/JNEUROSCI.3203-12.2012

Nozaradan, S., Peretz, I., & Mouraux, A. (2012b). Steady-state evoked potentials as an index of multisensory temporal binding. *NeuroImage*, *60*(1), 21–28. https://doi.org/10.1016/j.neuroimage.2011.11.065

Nozaradan, S., Schönwiesner, M., Caron-Desrochers, L., & Lehmann, A. (2016). Enhanced brainstem and cortical encoding of sound during synchronized movement. *NeuroImage*, *142*, 231–240. https://doi.org/10.1016/j.neuroimage.2016.07.015

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

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

Shen, L., Lu, X., Wang, Y., & Jiang, Y. (2023). Audiovisual correspondence facilitates the visual search for biological motion. *Psychonomic Bulletin & Review*, *30*(6), 2272–2281. https://doi.org/10.3758/s13423-023-02308-z

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

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809–813. https://doi.org/10.1073/pnas.0707021105

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *Journal of Neuroscience*, *25*(28), 6499–6508. https://doi.org/10.1523/JNEUROSCI.5095-04.2005

Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, *198*(2), 113–126. https://doi.org/10.1007/s00221-009-1880-8

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., Kurela, L. R., Siemann, J. K., James, T. W., & Wallace, M. T. (2014). Identifying and Quantifying Multisensory Integration: A Tutorial Review. *Brain Topography*, *27*(6), 707–730. https://doi.org/10.1007/s10548-014-0365-7

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), 14,1-11. https://doi.org/10.1167/10.12.14

Thomas, J. P., & Shiffrar, M. (2013). Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony. *Journal of Vision*, *13*(14), 8,1-13. https://doi.org/10.1167/13.14.8

Troje, N. F., & Westhoff, C. (2006). The Inversion Effect in Biological Motion Perception: Evidence for a "Life Detector"? *Current Biology*, *16*(8), 821–824. https://doi.org/10.1016/j.cub.2006.03.022

Ujiie, Y., Asai, T., & Wakabayashi, A. (2015). The relationship between level of autistic traits and local bias in the context of the McGurk effect. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00891

Vallortigara, G., & Regolin, L. (2006). Gravity bias in the interpretation of biological motion by inexperienced chicks. *Current Biology*, *16*(8), R279–R280. https://doi.org/10.1016/j.cub.2006.03.052

van der Zwan, R., MacHatch, C., Kozlowski, D., Troje, N. F., Blanke, O., & Anna, B. (2009). Gender bending: Auditory cues affect visual judgements of gender in biological motion displays. *Experimental Brain Research*, *198*(2–3), 373–382. https://doi.org/10.1007/s00221-009-1800-y

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 625–629. https://doi.org/10.3758/BF03206542

Wang, L., & Jiang, Y. (2012). Life motion signals lengthen perceived temporal duration.

Proceedings of the National Academy of Sciences, *109*(11), E673-677. https://doi.org/10.1073/pnas.1115515109

Wang, L., Yang, X., Shi, J., & Jiang, Y. (2014). The feet have it: Local biological motion cues trigger reflexive attentional orienting in the brain. *NeuroImage*, *84*, 217–224. https://doi.org/10.1016/j.neuroimage.2013.08.041

Wang, Y., Zhang, X., Wang, C., Huang, W., Xu, Q., Liu, D., Zhou, W., Chen, S., & Jiang, Y. (2022). Modulation of biological motion perception in humans by gravity. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-30347-y

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory Interactions along Lateral Temporal Regions Evoked by Audiovisual Speech. *Cerebral Cortex*, *13*(10), 1034–1043. https://doi.org/10.1093/cercor/13.10.1034

Wuerger, S. M., Crocker-Buque, A., & Meyer, G. F. (2012). Evidence for auditory-visual processing specific to biological motion. *Seeing and Perceiving*, *25*(1), 15–28. https://doi.org/10.1163/187847611X620892

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

Figure

Fig.1

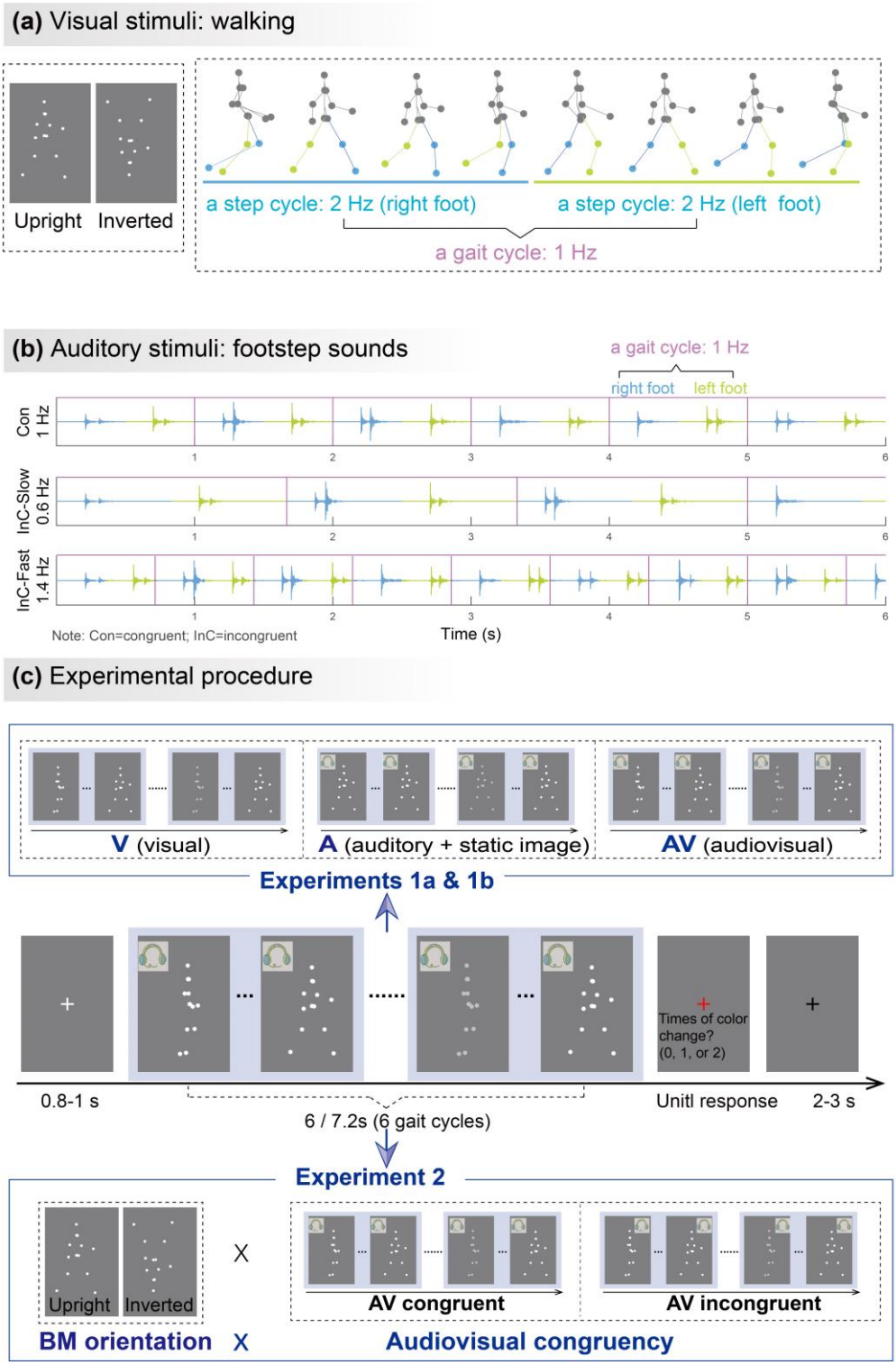
(a) Visual stimuli: walking
Upright
Inverted
a step cycle: 2 Hz (right foot)
a step cycle: 2 Hz (left foot)
a gait cycle: 1 Hz
(b) Auditory stimuli: footstep sounds
a gait cycle: 1 Hz
right foot
left foot
Con 1 Hz
InC-Slow 0.6 Hz
InC-Fast 1.4 Hz
Note: Con=congruent; InC=incongruent
Time (s)
(c) Experimental procedure
V (visual)
A (auditory + static image)
AV (audiovisual)
Experiments 1a & 1b
Times of color change? (0, 1, or 2)
0.8-1 s
6 / 7.2s (6 gait cycles)
Unitl response
2-3 s
Experiment 2
Upright
Inverted
X
AV congruent
AV incongruent
BM orientation
X
Audiovisual congruency

Fig. 1 Illustrations of audiovisual stimuli and experimental procedures. The illustration was based on stimuli with a gait-cycle frequency of 1 Hz. **(a) Visual stimuli.** The left panel depicts the static schematic of upright and inverted point-light walkers. The right panel shows the keyframes from a gait cycle of the BM sequence. The colors of dots and lines between dots are for illustration only and are not shown in the experiments. **(b) Auditory stimuli.** The auditory sequences contain periodic impulses of footstep sounds whose peak amplitudes occur around the points when the foot strikes the ground. The duration of two successive impulses defines the gait cycle of footstep sounds, which is temporally congruent (Con) or incongruent (InC) with the visual stimuli. **(c) Experimental procedure and design**. The color of the visual stimuli changed one or two times within 6 s in the catch trials but did not change in the experimental trials. Participants were required to report the number of changes when the point-light stimulus was replaced by a red fixation. In Experiment 1, participants viewed rhythmic point-light walkers or/and listened to the corresponding footstep sounds under visual (V), auditory (A), and audiovisual (AV) conditions. The visual stimulus was the BM sequence in the V and AV conditions but a static frame from the sequence in the A condition. Experiment 2 included only the AV condition with different stimulus orientations (upright vs. inverted) and audiovisual congruency (congruent vs. incongruent).

Fig.2

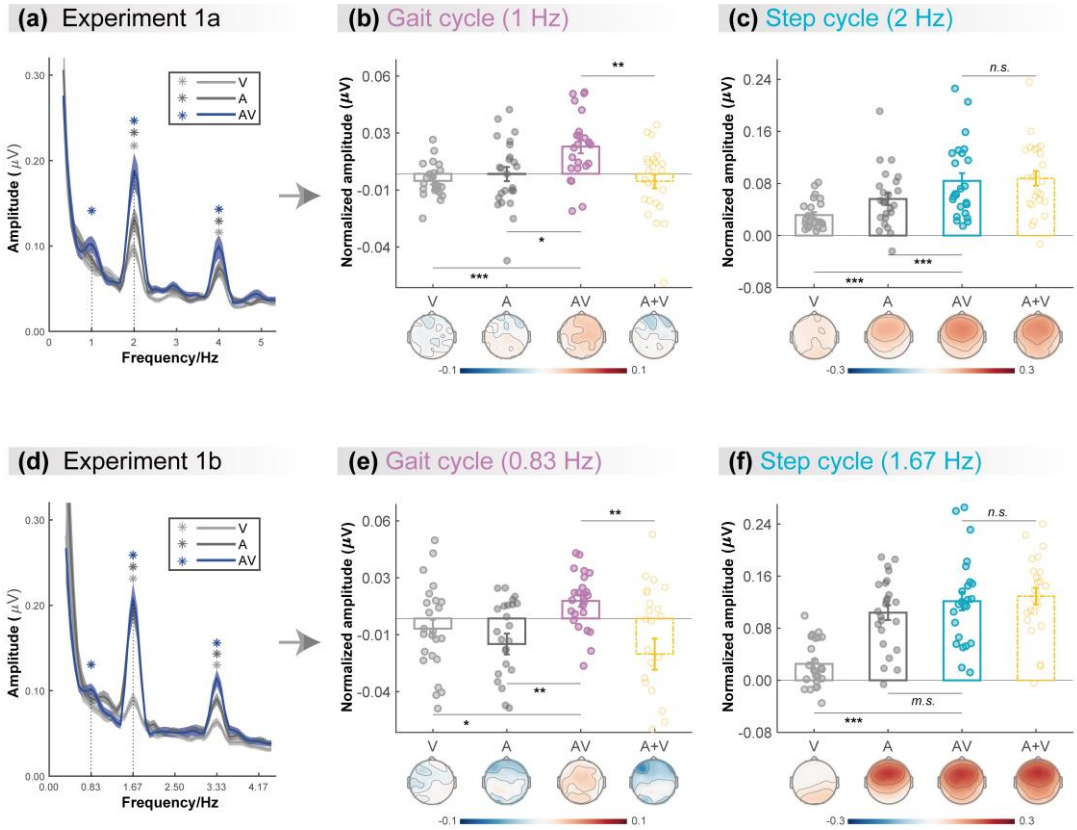

Fig. 2 Cortical tracking of visual (V), auditory (A), and audiovisual (AV) BM signals at gait-cycle and step-cycle frequencies. (a) & (d) The amplitude spectra of EEG response in three conditions in Experiment 1a and Experiment 1b, respectively. The solid lines show the grand average amplitude over all electrodes and subjects. The shaded regions depict standard errors of the group mean. Asterisks indicate significant spectra peaks (one-sample t-test against zero; $p < .05$, FDR corrected). **(b) & (e)** The normalized amplitude at gait-cycle frequency in the AV condition exceeded the arithmetical sum of those in V and A conditions (AV > A+V), **(c) & (f)** but the normalized amplitude at step-cycle frequency in the AV condition was comparable to the sum of V and A (AV = A+V). Colored dots represent individual data in each condition. Error bars represent ±1 standard error of means. *: $p < .05$; **: $p < .01$; ***: $p < .001$; *m.s.*: $.05 < p < .10$; *n.s.*: $p > .05$.

Fig.3

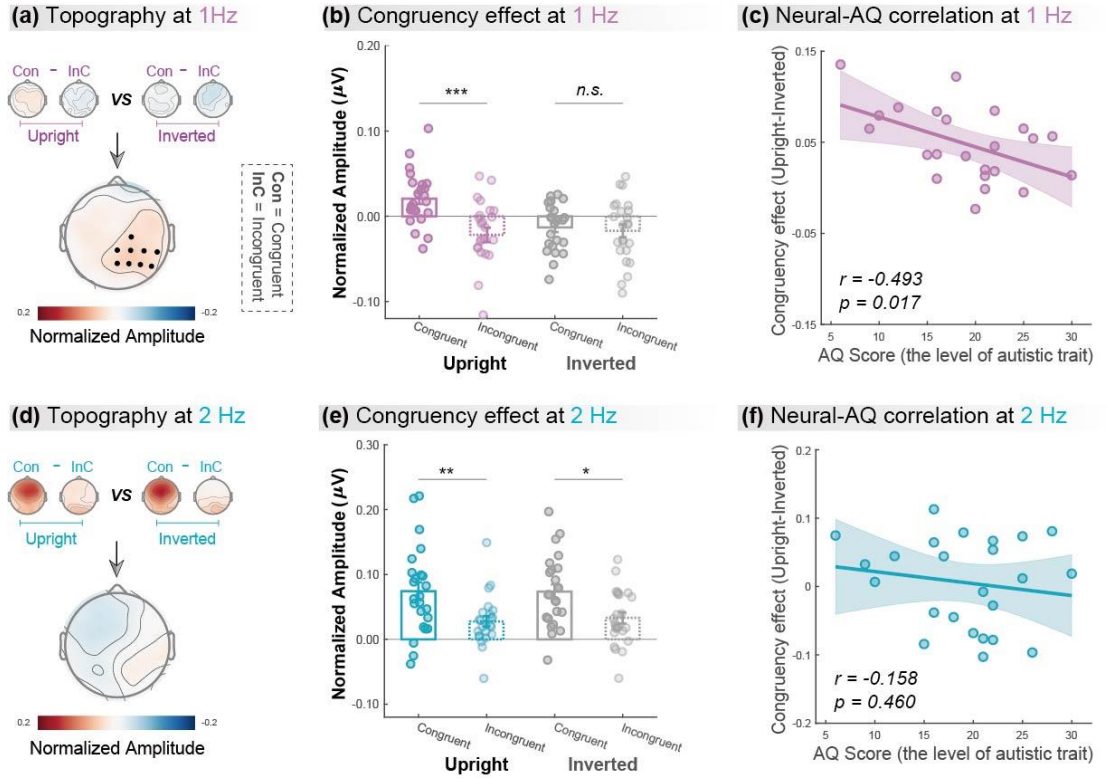

Fig. 3 Cortical tracking at gait-cycle rather than step-cycle frequency contributes to the BM-specific audiovisual integration (AVI) effect. The lower panels in **(a)** and **(d)** depict the topographic maps of the BM-specific AVI effect, measured by the difference of congruency effect between the Upright and Inverted conditions at 1 Hz and 2 Hz, respectively. A significantly enhanced congruency effect in the upright condition relative to the inverted condition was observed at 1 Hz (marked by black dots) but not at 2 Hz. The amplitude at these significant electrodes was averaged to quantify the congruency effect for the upright and inverted conditions at 1 Hz **(b)** and 2 Hz **(e)**. Error bars represent ±1 standard error of means. *: $p < .05$; **: $p < .01$; ***: $p < .001$; *n.s.*: $p > .05$. Individuals' autistic traits correlated with the BM-specific AVI effect at 1 Hz **(c)** but not at 2 Hz **(f)**. Shaded regions indicate the 95% confidence intervals.

Supplementary Information

Results on other peaks in Experiment 1

As shown in **Fig. 2a&d**, the audiovisual BM signals induced significant amplitude peaks at 1f (1/0.83 Hz), 2f (2/1.67 Hz), and 4f (4/3.33 Hz) relative to the gait cycle frequency (*p*s < .001; FDR corrected). To further test the functional roles of the neural activity at different frequencies, we analyzed the audiovisual integration modes at each frequency, by comparing the neural responses in the AV condition with the sum of those in the A and V conditions. Given that Experiments 1a & 1b yielded similar results, we collapsed the data and presented the results as follows.

As shown in **Fig.S1**, at 4f, the amplitude of neural responses showed significant peaks in all three conditions (V: t (47) = 6.869, p < 0.001; A: t (47) = 7.938, p < .001; AV: t (47) = 8.303, p < 0.001). Moreover, the amplitude in the AV condition was larger than that in the V condition (t (47) = 4.855, p < .001, Cohen's d = 0.701;) and the A condition (t (47) = 3.080, p = .003, Cohen's d = 0.445), respectively, suggesting multisensory gains. In addition, the amplitude in the AV condition was comparable to the unisensory sum (t (47) = -1.049, p = .300, Cohen's d = -0.151), indicating linear audiovisual integration. These results were similar to those observed at 2f but different from those at 1f, as reported in the main text. Together, these results show a similar additive audiovisual integration mode at 2f and 4f and a super-additive integration mode only at 1f, suggesting that the cortical tracking effects at 2f and 4f may be functionally linked but independent of that at 1f.

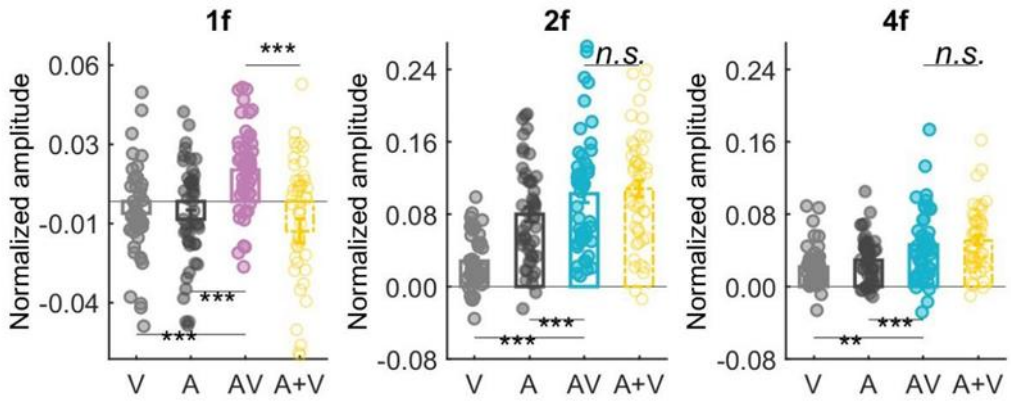

Fig. S1 Cortical tracking of audiovisual BM information at different frequencies. Colored dots represent individual data in each condition. Error bars represent ±1 standard error of means. *: $p < .05$; **: $p < .01$; ***: $p < .001$; n.s.: $p > .05$.

Control analysis of correlation in Experiment 2

The control analysis mainly aims to eliminate the potential bias due to electrode selection. As reported in the main text, both correlation analyses at 1 Hz and 2 Hz were performed based on electrodes in the significant cluster observed at 1 Hz because there was no significant cluster at 2 Hz (**Fig. 3a&d**, lower panel**)**. There is a possibility that these electrodes did not show a significant congruency effect at 2 Hz, either in the upright or the inverted condition, thus were not able to capture the correlation between the variance in neural responses and that in autistic traits. To rule out such a possibility, we conducted a control analysis based on electrodes showing a significant congruency effect at 2 Hz, for the upright (p = .004, cluster-based permutation test) and inverted (p = .002, cluster-based permutation test) conditions (**Fig. S2a**), respectively. As shown in **Fig.S2a**, both upright and inverted stimuli induced a significant congruency effect at 2 Hz (Upright: t (23) = 4.217, p < 0.001, Cohen's d = 0.861; Inverted: t (23) = 5.072, p < 0.001, Cohen's d = 1.035).

The difference of congruency effect between upright and inverted conditions is still not significant in the group level (t (23) = -0.689, *p* = 498, Cohen's d = -0.141), while it shows individual variance (SD = 0.079, range: [-0.173 0.153]) larger than that for the 1 Hz condition (SD = 0.041, range: [-0.023 0.135]), which allows us to identify a correlation if existing. Analysis of these data showed a non-significant correlation (**Fig. S2b**, r = -0.091, *p* = .674), similar to the results illustrated in **Fig. 3f**.

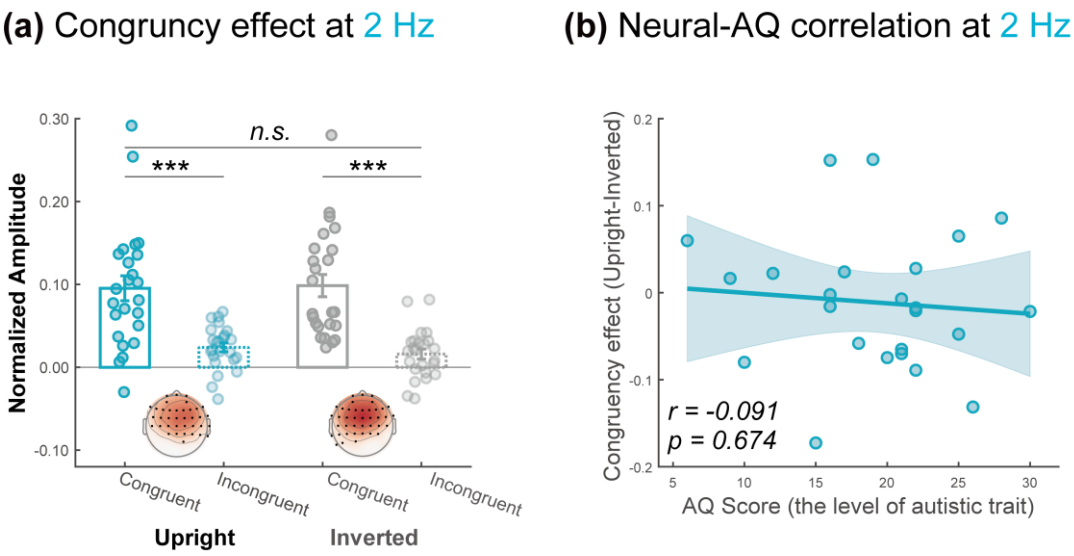

Fig. S2 Control analysis at step-cycle frequency. (a) The amplitude at the electrodes marked by solid black dots was averaged to quantify the cortical tracking effect under the upright and inverted conditions, respectively. The congruency effect was not significantly different between these conditions at the group level. **(b)** The individual congruency effect in the upright BM condition over the inverted condition was not significantly correlated with the AQ score.

Additional analysis in Experiment 2

To further examine the functional relevance between autistic traits and the BM-specific cortical tracking effect, we split the participants into high (above 20) and low (below or equal to 20) AQ groups by the median AQ score (20) of this sample. Similar to correlation analysis, one outlier, whose BM-specific audiovisual congruency effect

(Upright – Inverted) in neural responses at 1 Hz exceeds 3 SD from the group mean, was removed from the following analysis. As shown in **Fig.S3**, at 1 Hz, participants with low AQ showed a greater cortical tracking effect compared with high AQ participants (t (21) = 2.127, p = 0.045). At 2 Hz, low and high AQ participants showed comparable neural responses (t (22) = 0.946, p = 0.354). These results are in line with the correlation analysis, providing further support to the functional relevance between social cognition and cortical tracking of biological motion as well as its dissociation at the two temporal scales.

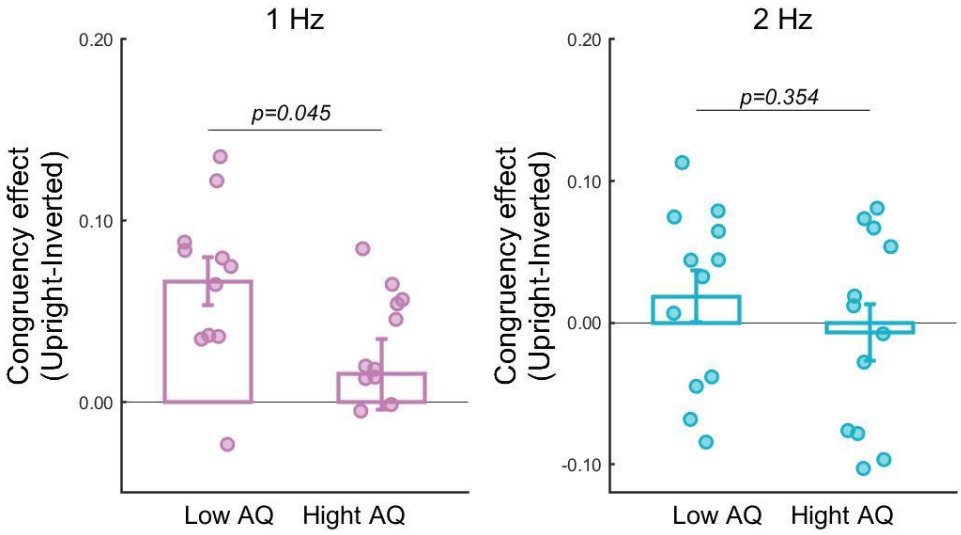

Fig. S3 The BM-specific cortical tracking effect for high and low AQ groups at 1 Hz and 2 Hz