Title page

Hierarchical cortical entrainment 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, CAS Center for Excellence in Brain Science and Intelligence Technology, 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 achieve efficient perception and develop adaptive responses. However, how human brains integrate these complex audiovisual cues based on their natural temporal correspondence remains unknown. Using electroencephalogram, we demonstrated that cortical oscillations entrained to hierarchical rhythmic structures in audiovisually congruent human walking movements and footstep sounds. Remarkably, the entrainment effects at different time scales exhibit distinct modes of multisensory integration, i.e., an additive integration effect at a basic-level integration window (step-cycle) and a super-additive multisensory enhancement at a higher-order temporal integration 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 hierarchical cortical entrainment orchestrates the processing of complex, rhythmic stimuli in natural contexts.

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

Introduction

The perception of biological motion (BM), the movements of living creatures, is a fundamental ability of the human visual system that is crucial in survival and social situations. 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 situations, e.g., one may simultaneously hear footstep sounds while seeing others walking. The integration of these visual and auditory BM cues based on congruency in stimulus contents or temporal relationships can facilitate the detection and discrimination of BM stimuli (Mendonça et al., 2011; Shen, Lu, Wang, et al., 2023; Thomas & Shiffrar, 2013; van der Zwan et al., 2009). Remarkably, such a cross-modal effect appears to engage an audiovisual integration (AVI) mechanism specific to BM, as the effect disappeared when the visual BM signals were deprived of characteristic kinematic cues but not low-level motion attributes through stimulus inversion (Brooks et al., 2007; Thomas & Shiffrar, 2010), and the temporal windows of perceptual audiovisual synchrony are different between BM and other motion stimuli with constant motion speed or gravity-incompatible accelerations (Arrighi et al., 2006; Saygin et al., 2008). Despite the behavioral evidence, the neural basis for the AVI of BM signals based on their natural multisensory correspondence remains largely unclear.

An intrinsic property of human movements (such as walking and running) is that

they are rhythmic and accompanied by frequency-congruent sounds. The AVI of such rhythmic stimuli may involve cortical entrainment, a process that the neural oscillations in cortical networks entrain to external rhythms and show increased activity or phase coherence at corresponding frequencies (Bauer et al., 2020; Lakatos et al., 2019). Studies based on simple or discrete stimuli have found that temporal congruency in auditory and visual rhythms significantly enhances the cortical tracking of rhythmic stimulations in both modalities (Nozaradan et al., 2012b). Unlike these stimuli, BM conveys complex hierarchical rhythmic structures that could be extracted from integration windows at different 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 entrainment to these hierarchical 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 rhythmic signals 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**). A greater cortical entrainment effect in the AV condition compared to each unisensory condition will indicate significant multisensory gains. Moreover, contrasting the multisensory response with the

summation of the unisensory responses serves to distinguish among sub-additive (AV < A+V), additive (AV = A+V), and super-additive (AV > A+V) modes of multisensory integration (see a review by Stevenson et al., 2014). Experiment 2 further 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 responses. Participants perceived the visual stimuli accompanied by temporally congruent or incongruent BM sounds. Comparing the congruency effect in neural responses between the upright and inverted conditions provides a unique opportunity to verify whether the AVI of BM involves a mechanism distinct from that underlies the AVI of non-BM information.

It is also worthy to note that the abilities to process BM information and integrate multisensory inputs vary across individuals and are diminished in populations with autism spectrum disorder (ASD) or even high autistic traits (Feldman et al., 2018; Pavlova, 2012; Wang et al., 2018). Specifically, ASD individuals showed reduced orienting to audiovisually synchronized BM stimuli (Klin et al., 2009), and such impairment in 10-month infancy can predict autism diagnosis at 3 years of age (Falck-Ytter et al., 2018). These findings suggest a possible link between compromised audiovisual BM processing ability and higher autistic-like traits, given that social cognitive deficits in ASD lie on a continuum extending from the clinical to

nonclinical populations with different levels of autistic traits (Baron-Cohen et al., 2001). Therefore, here we examined the potential relationship between participants' neural responses to synchronous audiovisual BM signals and their autistic traits in Experiment 2.

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. We were interested in two critical rhythmic structures in the walking motion sequence, i.e., the gait cycle and the step cycle (**Fig. 1a**). 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) = 6.964, 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 a significant peak at 4 Hz in all three conditions (ps < 0.001, FDR corrected), which might be the harmonic of 2 Hz (see the results of harmonics 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 cortical entrainment 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 entrainment to 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, see Supplementary Information for additional analysis). 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.

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 entrainment effects at the two temporal scales might be driven by functionally different mechanisms.

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

To further explore whether and how the cortical tracking of the rhythmic information contributes to the specialized audiovisual process 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. 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 the AVI process specific to BM information.

We calculated the audiovisual congruency effect for the upright (AVIupr) and the inverted (AVIinv) conditions, respectively. Then, we identified the clusters showing significantly different congruency effects between the upright and inverted conditions 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 a two-tailed paired t-test. The results showed that (**Fig. 3b**) the 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). The congruency effect in the upright condition was significantly larger than that in the inverted condition (t (23) = 3.099, p = 0.005, Cohen's d = 0.633).

In contrast, at 2 Hz, no cluster showed a significantly different congruency effect between the upright and inverted conditions (**Fig. 3d**). We then conducted further analysis on the averaged amplitude of the electrodes marked in **Fig.** 3a (lower panel). A two-tailed paired t-test 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). The congruency effect between the upright and inverted conditions was not different (t (23) = 0.434, p = 0.668, Cohen's d = 0.089), suggesting a comparable audiovisual congruency effect between two conditions at 2 Hz. Importantly, a three-way repeated-measures ANOVA with frequency (1 Hz vs. 2 Hz), orientation (upright vs. inverted), and audiovisual congruency (congruent vs. incongruent) as within-subject factors revealed a marginal significant three-way interaction (F (1,23) = 3.190, p = 0.087, η_p^2 = 0.122), further implying 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 (exceeded 3 SD), 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 electrodes within the clusters showing significant congruency effects at 2 Hz (see the control analysis in Supplementary Information for details).

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, neural oscillations of observers entrained to 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. Moreover, the strength of these cortical entrainment effects was enhanced under the audiovisual condition than in the visual-only or auditory-only condition, indicating multisensory gains in the cortical tracking of BM information (Experiments 1a & 1b).

Crucially, although the entrainment processes at both gait-cycle frequency and step-cycle frequency gain benefits from multisensory correspondence, the mechanisms underlying these two processes appear to be different. At step-cycle frequency, the cortical entrainment effect in the AV condition equals the additive sum of the unisensory conditions. Such linear integration might 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 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 entrainment to gait-cycle rather than step-cycle is specific to the AVI of BM. 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. These findings suggest that the cortical entrainment at step-cycle frequency reflects the integration of basic motion signals and corresponding sounds, while the cortical entrainment at gait-cycle frequency reflects the AVI of higher-level BM information. 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.

Besides the temporal dynamics of neural activity revealed by the cortical entrainment process, we found that the BM-specific AVI effect was associated with enhanced cortical tracking of gait cycles in the right temporoparietal electrodes. This finding likely relates to neural activity in 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 temporal resolution and spatial resolution, such as MEG.

In a broad sense, the current study deepens our understanding of the neural processing of audiovisual signals in natural stimuli with complex temporal structures. Cortical entrainment can track simple rhythmic stimuli like tone sequences or luminance-varying patches (C. Keitel et al., 2017; Yuan et al., 2021) as well as complex rhythmic structures in speech (Brookshire et al., 2017; Ding et al., 2016; Keitel et al., 2018) and BM (Shen, Lu, Yuan, et al., 2023). Beyond unisensory processing, cortical entrainment also plays a role in the multisensory processing of simple or discrete rhythmic signals generated by physical stimulation (Bauer et al.,

2021; Miller et al., 2013; Nozaradan et al., 2012b; Simon & Wallace, 2017). These findings may partially explain the AVI effect at 2 Hz for BM and non-BM stimuli in the current study. However, we found that the cortical tracking of the perceived higher-order rhythmic structure based on 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. Similar to BM, other natural rhythmic stimuli, like auditory speech, also convey hierarchical structures that can entrain neural oscillations at different temporal scales (Ding et al., 2016; Keitel et al., 2018). Previous studies have observed the AVI of speech at theta band (4-6 Hz), a temporal scale that corresponds to the rate of syllables (Crosse et al., 2015), and that the asynchrony detection of prosodic fluctuation in audiovisual speech linked with delta oscillations (~1-3 Hz) (Biau et al., 2022). These findings raise a possibility that the AVI of speech also occurs at multiple temporal scales and that the multi-scale entrainment effects play different roles in speech perception. Further investigation into these issues 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 cortical tracking and integration 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 entrainment to the temporal correspondence 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). This point-light walker looked like walking on a treadmill and did not translate on the screen. 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 entrainment 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 (Maris & Oostenveld, 2007) 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 entrainment 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) (Maris & Oostenveld, 2007). 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 2021ZD0203800 and 2021ZD0204200), the National Natural Science Foundation of China (Nos. 32171059 and 31830037), the Interdisciplinary Innovation Team (JCTD-2021-06), the Youth Innovation Promotion Association of the Chinese Academy of Sciences, and the Fundamental Research Funds for the Central Universities.

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://osf.io/6f7t4/.

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

Bauer, A.-K. R., van Ede, F., Quinn, A. J., & Nobre, A. C. (2021). Rhythmic Modulation of Visual Perception by Continuous Rhythmic Auditory Stimulation. *The Journal of Neuroscience*, *41*(33), 7065–7075. https://doi.org/10.1523/JNEUROSCI.2980-20.2021

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

Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences*, *114*(24), 6352–6357. https://doi.org/10.1073/pnas.1620350114

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

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

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, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, *16*(3), e2004473. https://doi.org/10.1371/journal.pbio.2004473

Keitel, C., Thut, G., & Gross, J. (2017). Visual cortex responses reflect temporal structure of continuous quasi-rhythmic sensory stimulation. *NeuroImage*, *146*, 58–70. https://doi.org/10.1016/j.neuroimage.2016.11.043

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

Lakatos, P., Gross, J., & Thut, G. (2019). A New Unifying Account of the Roles of Neuronal Entrainment. *Current Biology*, *29*(18), R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

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

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

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

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

Miller, J. E., Carlson, L. A., & McAuley, J. D. (2013). When What You Hear Influences When You See: Listening to an Auditory Rhythm Influences the Temporal Allocation of Visual Attention. *Psychological Science*, *24*(1), 11–18. https://doi.org/10.1177/0956797612446707

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

Pavlova, M. A. (2012). Biological Motion Processing as a Hallmark of Social Cognition. *Cerebral Cortex*, *22*(5), 981–995. https://doi.org/10.1093/cercor/bhr156

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

Simon, D. M., & Wallace, M. T. (2017). Rhythmic Modulation of Entrained Auditory Oscillations by Visual Inputs. *Brain Topography*, *30*(5), 565–578. https://doi.org/10.1007/s10548-017-0560-4

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

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, Y., Wang, L., Xu, Q., Liu, D., Chen, L., Troje, N. F., He, S., & Jiang, Y. (2018). Heritable aspects of biological motion perception and its covariation with autistic traits. *Proceedings of the National Academy of Sciences*, *115*(8), 1937–1942. https://doi.org/10.1073/pnas.1714655115

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

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

Yuan, P., Hu, R., Zhang, X., Wang, Y., & Jiang, Y. (2021). Cortical entrainment to hierarchical contextual rhythms recomposes dynamic attending in visual perception. *eLife*, *10*, e65118. https://doi.org/10.7554/eLife.65118

Figure

Fig.1

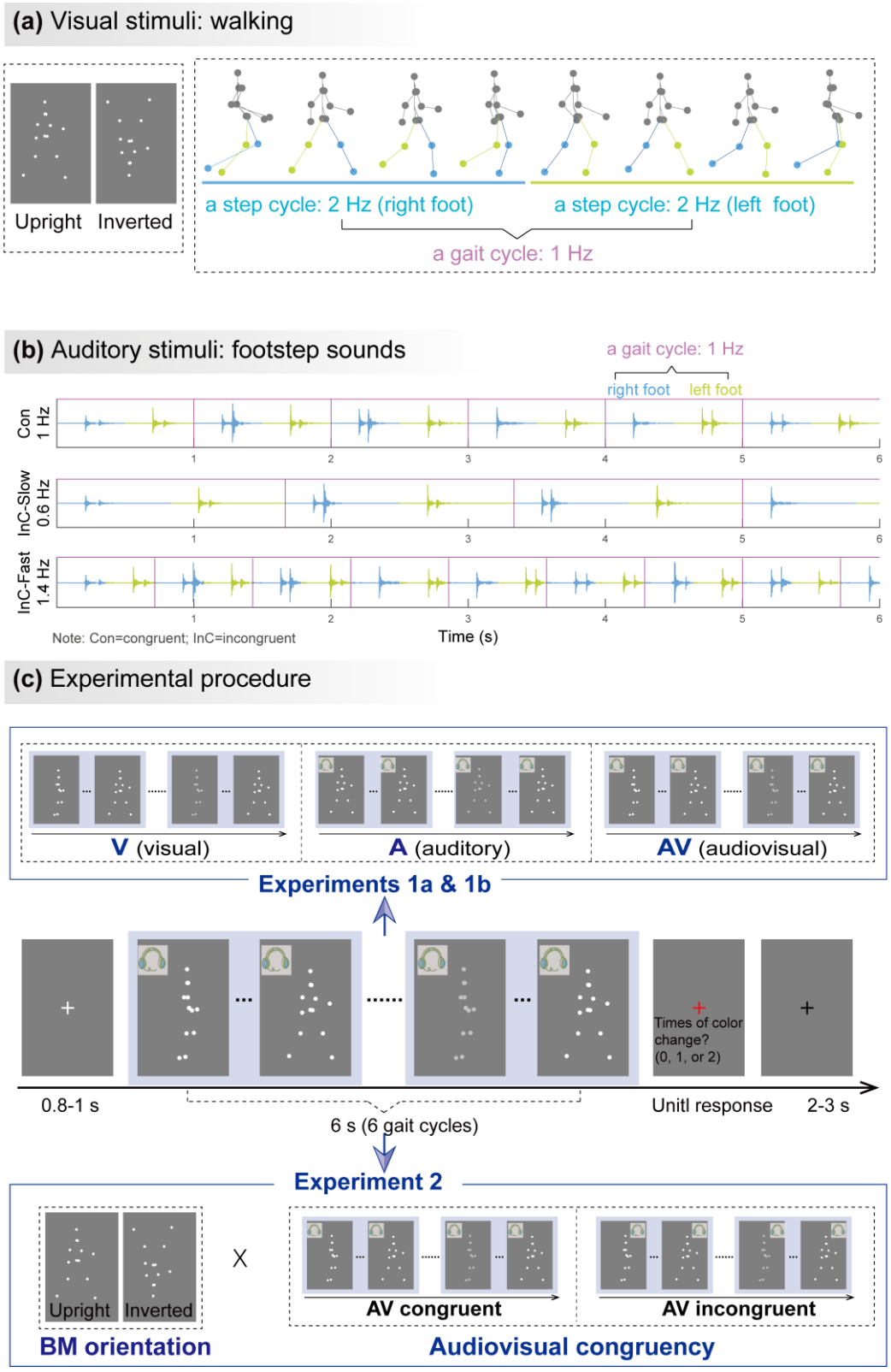
(a) Visual stimuli: walking
Upright
Inverted
a step cycle: 2 Hz (right foot)
a step cycle: 2 Hz (left foot)
a gait cycle: 1 Hz
(b) Auditory stimuli: footstep sounds
a gait cycle: 1 Hz
right foot
left foot
Con 1 Hz
InC-Slow 0.6 Hz
InC-Fast 1.4 Hz
Note: Con=congruent; InC=incongruent
Time (s)
(c) Experimental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

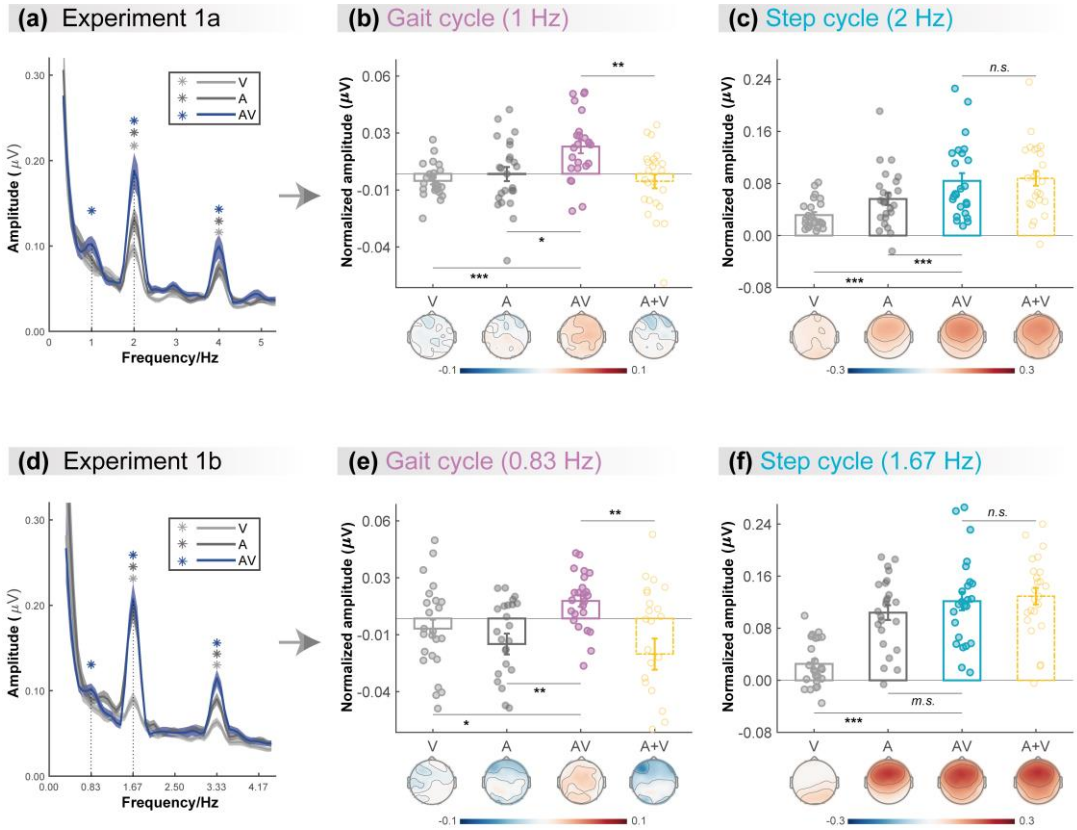

Fig. 2 Cortical tracking of visual (V), auditory (A), and audiovisual (AV) BM signals at gait-cycle and step-cycle frequencies. (a) & (d) The amplitude spectra of EEG response in three conditions in Experiment 1a and Experiment 1b, respectively. The solid lines show the grand average amplitude over all 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

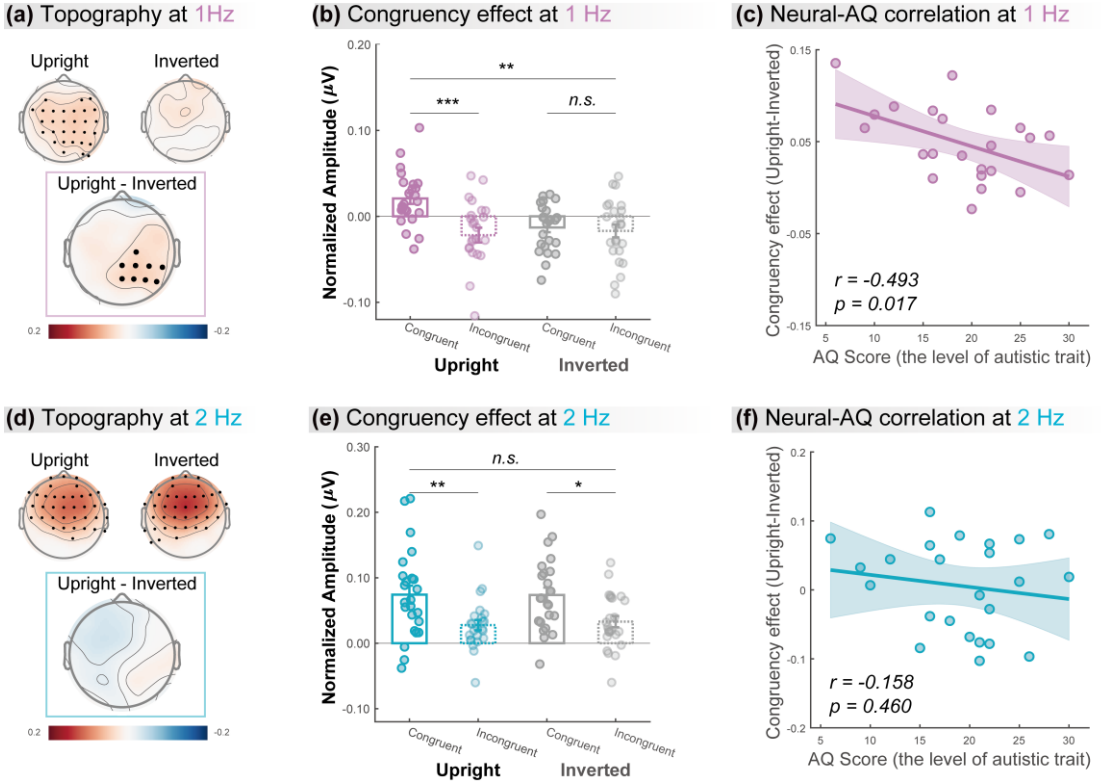

Fig. 3 Cortical tracking at gait-cycle rather than step-cycle frequency contributes to BM-specific AVI effect. (a) & (d) Topographic maps for the congruency effect at 1Hz and 2 Hz for each condition: Upright, Inverted, and for the difference between these conditions (Upright-Inverted), respectively. The black dots indicate the electrodes showing a significant congruency effect (upper panels) or a significantly enhanced congruency effect in the upright condition relative to the inverted condition (lower panel). Then, the amplitude at the electrodes shown in the lower panel of (a) were averaged to quantify the cortical entrainment effect at 1 Hz **(b)** and 2 Hz **(e)**. Error bars represent ±1 standard error of means. Individuals' autistic traits correlated with the BM-specific AVI at 1 Hz **(c)** but not 2 Hz **(f)**. Shaded regions indicate the 95% confidence intervals.

Supplementary Information

Results on harmonics in Experiment 1

As shown in Fig. 1a&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. No significant peak was observed at 3f (3/2.50 Hz) and 5f (5/4.17 Hz). Theoretically, 2f can be the harmonic component of f, and 4f can be the 4th harmonic of 1f and the 2nd harmonic of 2f (Norcia et al., 2015). If the fundamental oscillations and harmonic oscillations are generated via the same or tightly linked mechanisms (Abeysuriya et al., 2014), one may expect to observe similar patterns of results at the two frequencies. We conducted additional analyses to examine this issue. Given that Experiments 1a & 1b yielded similar results, we collapsed the data and presented the results as follows.

To explore the functional relationship 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. Results show that the integration mode at 1f is different from all others, while a similar additive audiovisual integration mode is observed at 2f and 4f (**Fig. S1a**, also see the results session in the main text for the detailed results at 1f and 2f). At 4f, the amplitude of neural responses showed significant peaks in all three conditions (V: t (47) = 6.869, p < .001; A: t (47) = 7.938, p < .001; AV: t (47) = 8.303, p < .001; FDR corrected). 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. There were no significant multisensory gains at 3f or 5f (**Fig. S1b**).

These results indicate that the response at 4f might be the harmonic of 2f, which plays a similar role as 2f in the audiovisual integration of biological motion. However, the cortical entrainment effect at 2f is functionally independent of 1f and can not be fully explained by the harmonic relationship.

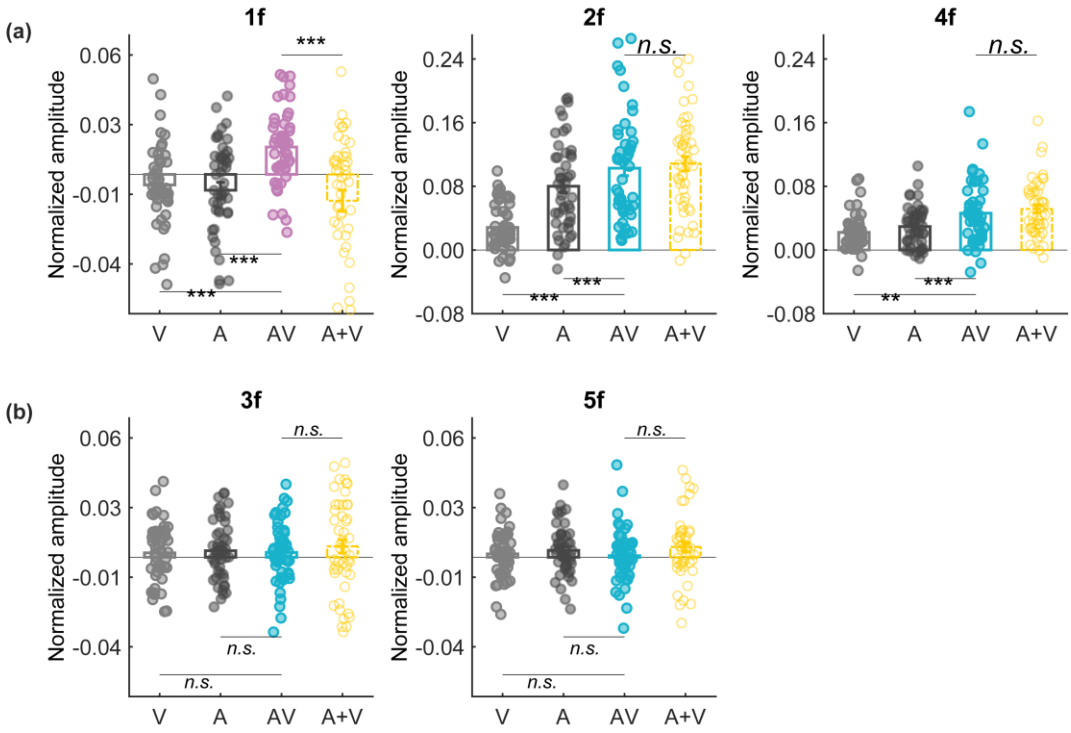

Fig. S1 Cortical tracking of audiovisual BM information at different frequencies

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. We further calculated the difference of congruency effects between these conditions. Note that while this index is not significant at the group level (t (23) = -0.689, p = 498), 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. 2f**.

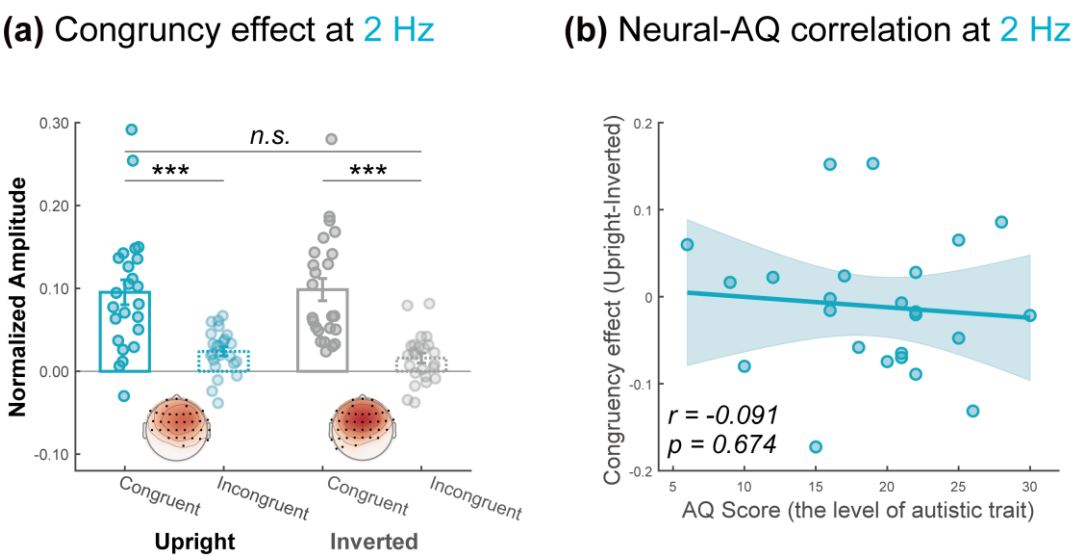

Fig. S2 Control analysis at step-cycle frequency. (a) The amplitude at the electrodes marked by solid black dots was averaged to quantify the cortical entrainment effect under the upright and inverted conditions, respectively. The congruency effect was not significantly different between these conditions 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.

References:

Abeysuriya, R. G., Rennie, C. J., Robinson, P. A., & Kim, J. W. (2014). Experimental observation of a theoretically predicted nonlinear sleep spindle harmonic in human EEG. *Clinical Neurophysiology*, *125*(10), 2016–2023. https://doi.org/10.1016/j.clinph.2014.01.025

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4,1-46. https://doi.org/10.1167/15.6.4