Research article

# Perceptual coupling in human dyads: Kinematics does not affect interpersonal synchronization

Mattia Rosso [*], Canan Nuran Gener, Bart Moens, Pieter-Jan Maes, Marc Leman

*IPEM - Institute for Systematic Musicology, Ghent University, Ghent, Flanders, 9000, Belgium*

A B S T R A C T

The minimal, essential condition for individuals to interact is that they exchange information via at least one sensory channel. Once informational coupling is established, it enables basic forms of coordinated behavior to spontaneously emerge from the interaction. Our previous study revealed different coordination dynamics in dyads engaged in a joint finger-tapping task based on visual versus auditory coupling. This observation led us to propose the 'modality-dependent hypothesis', which posits that coordination dynamics are influenced by the sensory modality mediating informational coupling. However, recognizing that different modalities have inherent differences in accessing spatiotemporal features of perceived movement, we formulated the alternative 'kinematic hypothesis'. This hypothesis posits that differences in dynamics would vanish given equivalent kinematic information across modalities. The study involved forty (N = 40) participants, grouped into twenty (N = 20) dyads, who engaged in a joint finger-tapping task. This task was conducted under varying conditions of visual and auditory coupling, with manipulations in the access to kinematic information, categorized as discrete and continuous. Contrary to our initial predictions, the results strongly supported the 'modality-dependent hypothesis'. We observed that visual and auditory coupling consistently yielded distinct attractor dynamics, regardless of the access to kinematic information. Furthermore, all conditions of auditory coupling resulted in higher levels of synchronization than their visual counterparts. These findings suggest that the differences in interpersonal synchronization are predominantly influenced by the sensory modality, rather than the continuity of kinematic information. Our study highlights the significance of sensorimotor interactions in interpersonal synchronization and addresses the potential of sonification strategies in supporting motor training and rehabilitation.

## 1. Introduction

In order to move and interact with their environments, humans have to process and integrate sensory inputs from different streams of information [1]. When the source is another human, the information encoded in their movement has the potential to guide mutual adaptation of behaviors and establish minimal forms of interaction [2]. It is self-evident that the processes of accessing and decoding information from perceived rhythmic behaviors differ across sensory modalities. When we listen to someone walking, clapping or drumming, we typically hear the footstep, the clap or the hit as discrete perceptual outcomes. In the auditory domain, this class of movements tend to be characterized by naturally mute phases [3]. We are 'blind' to the moment-by-moment course of their execution … until we observe how it evolves over time. Once visual contact is established, visual coupling provides individuals with continuous access to kinematic and geometric properties of an action [4], defined in more general terms as its informational observables [5].

Information is foundational to feed the action-perception loops underlying dyadic interactions [6], from the most basic level of spontaneous entrainment to more complex forms of planned joint action [7]. Within any biological system across different scales, informational coupling is the minimal condition for individual components to behave as coordinated units. In the case of a dyadic system constituted by two coupled individuals having independent nervous systems, coupling is necessarily based on perception via

---

* Corresponding author.
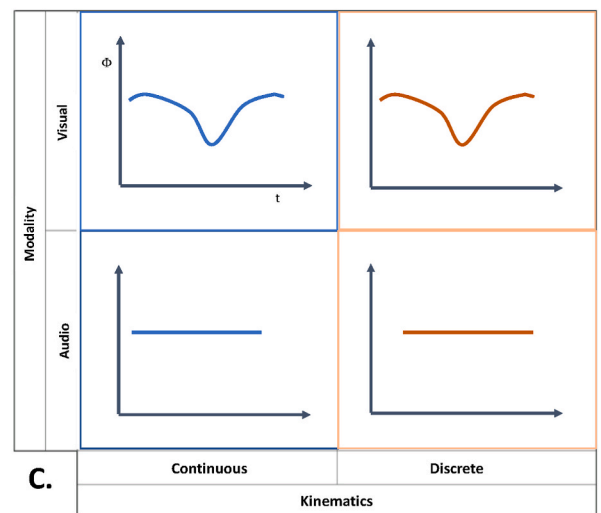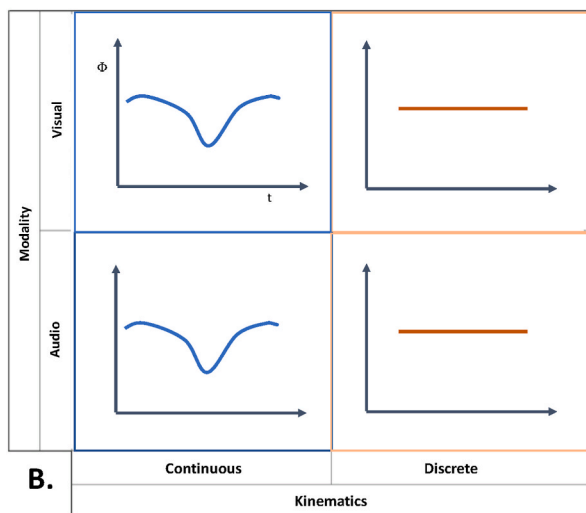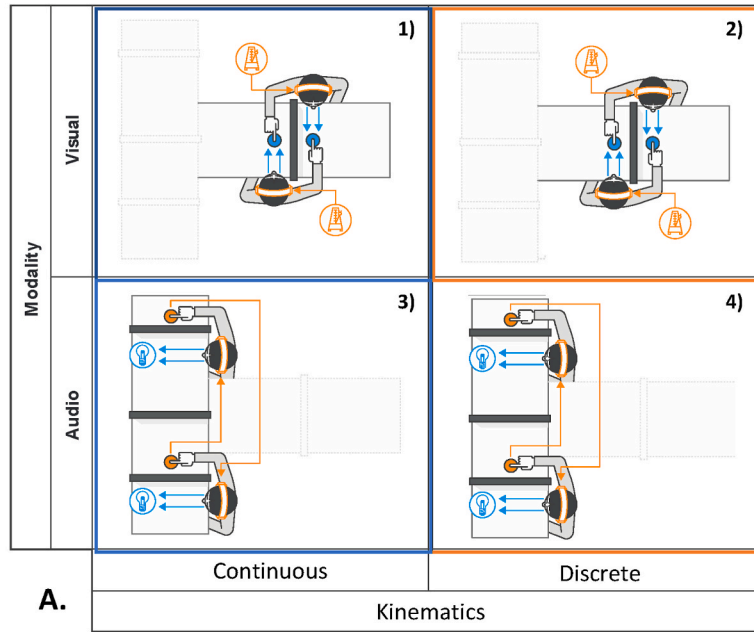*E-mail address:* mattia.rosso@ugent.be (M. Rosso).

**Fig. 1.** *Experimental design and hypotheses. A) Experimental design.* The study was designed in a Modality (Visual, Auditory) x Kinematics (Continuous, Discrete) factorial structure. The design resulted in the following experimental conditions. **1.** *Visually Coupled – Continuous*: participants tapped along with an auditory metronome, while looking at the partner's hand tapping. The hand of the partner was fully visible, while the view of their own hand was hidden by a component of the screen. **2.** *Visually Coupled – Discrete*: participants tapped along with an auditory metronome, while looking at the partner's hand tapping. The hand of the partner was hidden by a sliding window, adjusted so that only the fingertip was visible at the time of impact. The view of their own hand was hidden by a component of the screen. **3.** *Auditorily Coupled – Continuous*: participants tapped along with a flickering LED, while hearing the sonification of the partner's tapping. The sonification consisted of a continuous tone, modulated in frequency by the distance between the fingertip and the pad. A discrete 'tick' sound was triggered at the time of impact. The view of their own hand was hidden by a component of the screen. **4.** *Auditorily Coupled – Discrete*: participants tapped along with a flickering LED, while hearing the sonification of the partner's tapping. The sonification consisted of a discrete 'tick' sound triggered at the time of impact. The view of their own hand was hidden by a component of the screen. *B) Kinematic hypothesis.* This figure illustrates the expected outcomes of the joint finger-tapping task based on the 'kinematic hypothesis', wherein the coordination patterns are explained by the access to physical properties of motion (Continuous/Discrete) rather than by the sensory modality (Visual/Auditory). *C) Modality-dependent hypothesis.* This figure illustrates the results predicted by the complementary 'modality-dependent' hypothesis, in which coordination patterns are determined by the sensory modality (Visual/Auditory) rather than the physical properties of motion (Continuous/Discrete).The $\Phi$ symbol on the vertical axis indicates an order parameter, or the level of coupling in the system, whereas the $t$ on the horizontal axis indicates time. The depicted curves are for illustration purposes; intermediate scenarios may emerge from interaction effects across the factors.

one or multiple sensory channels. Once this condition is met, it sets the stage for the emergence of a dyadic control structure, formed over the informational structure of the system [4,5]. In other words, perception enables the coordination between two individuals, and the dynamic configuration of emerging behavioral patterns depends on the specifics of the perceptual information available in the interaction. Understanding how humans access such information, and how they guide their actions accordingly, is paramount for an ecological perspective on action and perception [8].

In a previous study, we reported different attractor dynamics in dyads engaged in a joint finger-tapping task, depending on whether participants could see (visual coupling) or hear (auditory coupling) the actions of their partner [9]. An attractor is the dynamical entity underlying entrainment, a basic control structure [4] which organizes the system's behavior in coordinated states. In this and other works [9–11], the *drifting metronomes* paradigm for dyadic entrainment was used to drive the dyad through a space of coordinative states and detect the presence of attractor points. This paradigm consisted of a joint-finger tapping task where each member of the dyad was assigned a metronome and instructed to synchronize their finger taps with it. The two metronomes were set to start in phase, but a minimal gap in their frequencies resulted in a linear de-phasing pattern: step after step, their relative phase increased from 0 to $\pi$ and subsequently decreased from $\pi$ to 0 before the cycle repeated. Although the participants were instructed to ignore the partner while maintaining synchronization with the assigned metronome, dyadic entrainment occurred spontaneously in both sensory modalities. However, only in the case of visual coupling did we observe emergent coordination patterns consistently shaped by an attractor landscape [12,13]. More precisely, over the space explored by the drifting metronomes, the in-phase point $(0,2\pi)$ stabilized the interaction in recurrent synchronized states, promoting a cooperation process in the dyadic system. The anti-phase point $(\pi)$ facilitated instead decoupled independent trajectories, promoting a competition process. Conversely, in the case of auditory coupling, the cooperation process was dominant but spread over the whole attractor landscape, with no significant influence of clear attractor points, and higher inter-dyad variability in terms of global entrainment.

Whilst our findings led us to conclude that attractor dynamics depend on the sensory modality mediating the coupling, we put forward the idea that the differences would attenuate or disappear if we could induce cross-modal correspondence between streams of information across visual and auditory couplings [1,14–16], therefore balancing the access to movement kinematics. The observation that in everyday environments visual rhythms tend to be continuous and auditory rhythms tend to be discrete [17] biased the way we implemented the sonification of participants' finger-taps in our previous study, so that they were presented to the partner as discrete sounds at the moment of impact [9]. Conversely, visual coupling was naturally continuous, and therefore not directly comparable to the auditory counterpart in terms of stimulus property. This limitation made it impossible to disentangle the contribution of the physical property of the stimulus (continuous/discrete) from the perceptual system accessing such information (visual/auditory). We argue that these constitute two distinct dimensions of informational coupling, which require a balanced factorial design to be independently manipulated (see Fig. 1).

While under ecological conditions the visual system exhibits preferential access to the observables of an action, continuously available in the optical structures it generates [4], sonification can be used as a means to convey kinematic information in the auditory domain [18–20]. Auditorily coupling can be therefore artificially augmented by expanding motion acoustics to the naturally mute phases of movement [3]. For finger-tapping, this would imply filling the mute inter-tap intervals by mapping movement parameters such as velocity, acceleration, or position, onto sonic variables such as pitch, amplitude, or spatial location. Auditory presentation of continuous rhythmic cues has been shown to affect movement control in sensorimotor synchronization tasks, better guiding online trajectory, enhancing control between endpoints, and improving consistency of motor timing. This same research suggests that sensorimotor synchronization is subserved by different processes, depending on whether the synchronizer has to internally self-time the intervals or can rely on extrinsic information [21]. When directly comparing coordination with auditory and visual environmental rhythms, differences tend to vanish when they are presented as continuous [17].

Translating these principles to the interpersonal domain, we expected to replicate the coordination patterns observed under visual coupling when continuous auditory access to the partner's kinematics was provided. Conversely, in a condition of discrete visual coupling, we anticipated that occluding the kinematic trajectory of the finger-tapping would result in a breakdown of the coordination pattern, given that participants would lose the advantage characterizing the visual modality [4]. Our prediction is substantiated by the main current theoretical accounts for interpersonal synchronization [22]. In a predictive coding framework, where attractor states result from the minimization of the mismatch between internal models [23], sonification would increase the temporal resolution of the model update by allowing regular and fine-grained sampling of the partner's movement cycles. In a dynamical systems framework, mapping kinematic information onto sound [24] would enable continuous coupling to drive the participants' behavior, so that synchronization can be achieved without explicitly computing the partner's inter-tap intervals [25]. We refer to this scenario, wherein attractor dynamics are explained by the access to the physical properties of motion, as the '*kinematic hypothesis*'.

Alternatively, we may observe a scenario wherein attractor dynamics are explained by sensory modality alone. Vision and audition evolved phylogenetically and ontogenetically with their specifics strategies for sampling environmental rhythms, and developed specific interface with motor timing system of the perceiver [26]. These being the two main perceptual systems mediating informational coupling, we may expect that they are optimized for subserving different aspects of dyadic interactions. In this case, these differences would be reflected in different coordination dynamics across modalities, regardless of the continuous or discrete nature of the information available to the perceiver. We refer to this alternative scenario as the '*modality-dependent hypothesis*'.

The aim of the present work is to answer the question of whether interpersonal synchronization relies on a supramodal processing of kinematic information ('kinematic hypothesis') or whether it depends on processing specific to the sensory modality and its unique interface with the motor system ('modality-dependent hypothesis'). To differentiate between these two scenarios, we experimentally disentangled the contribution of the specific perceptual system processing the information from the physical properties of the information itself. Fig. 1 provides a visual representation of the experimental design and the results we expected based the two

hypotheses. Whilst these are presented as the main expected outcomes, the design leaves room for intermediate scenarios and interactions between the two dimensions.

## 2. Results

Joint recurrence quantification analysis (JRQA) [27] was performed as described in Rosso et al. (2021b, 2023) [9,11] in order to quantify the degree of temporal coordination between the partners throughout the joint finger-tapping task [28]. The resulting recurrence score yielded a proxy measure of coupling strength between the partners, quantifying instances of coordinated behavior [27]. By tracking its variations over the course of the drifting metronomes' cycle, we were able to estimate the attractor landscape underlying the interaction [12,13]: the higher the recurrence score, the stronger the attraction to jointly revisit a coordinative state at a given value of the metronome's relative phase. The recurrence score's timeseries were used as response variable in our statistical model and modeled using growth curve analysis [29] (see '*Methods – Statistical model*'). For each dyad, curves were calculated as the average of 10 metronomes' cycles from 0 to $2\pi$ in every experimental condition and modeled with 2nd order orthogonal polynomials of Time. The 2-level factors, Modality (Visual, Auditory) and Kinematics (Continuous, Discrete), were modeled as fixed effects. Not having uncoupled conditions in the experimental design, effects were assessed by directly contrasting coupled conditions across levels of Modality and Kinematics. Auditory and Discrete were chosen as baseline levels for statistical comparisons, respectively.

Against our prediction, these findings provide robust evidence in favor of the '*modality-dependent hypothesis*', decisively refuting the '*kinematic hypothesis*'. This is confirmed by the significant interactions between Modality and the linear (Estimate = $-1049.894$, SE = $339.775$, p = $0.002$) and quadratic (Estimate = $1230.757$, SE = $470.396$, p = $0.008$) terms of Time. Taken together, the effects indicate that visually coupled conditions resulted in the same asymmetrical parabolic modulation reported in Refs. [9,11]. The linear term of the model captures the shift in the vertex of the parabola, whereas the quadratic term captures the depth of the modulation. This effect was irrespective of levels of Kinematics. Furthermore, we reported a main effect of Modality (Estimate = $-532.994$, SE = $207.089$, p = $0.010$), indicating that auditory coupling resulted in higher recurrence overall, in spite of the absence of modulation over the attractor landscape. This effect held irrespective of the levels of Kinematics. Crucially, the absence of any other effect indicates that discrete or continuous access to the partner's movements do not differentiate across the observed attractor dynamics, nor across global levels of recurrence.

Fig. 2 clearly shows that the grand-average curves of the recurrence score tend to overlap within the same Modality, and are not
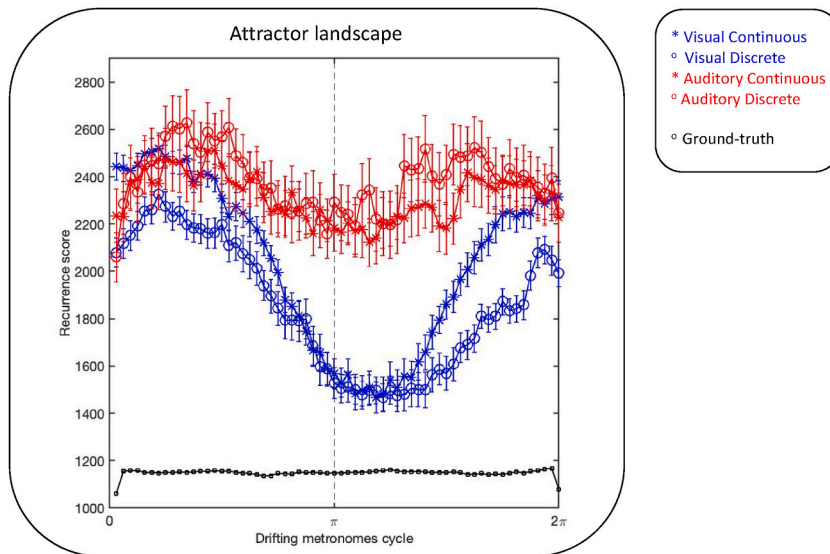


**Fig. 2.** *Attractor landscape.* The timeseries depicted in the figure represent the evolution of the joint recurrence score as a function of the drifting metronomes' cycle across experimental conditions. The grand average was computed over the entire sample of dyads (N = 19), with each dyad's timeseries calculated as the average of 10 consecutive cycles. The vertical bars represent the standard error of the mean (SEM) for each phase bin. For illustration and interpretation purposes, the black line in the plot shows the same analysis performed on the two metronomes' timeseries, providing the reference recurrence score expected by a deterministic decoupled system. A horizontal line lingering at the global minimum represents the expected pattern for two partners perfectly synchronizing with their assigned metronome without influencing each other. The black line exhibits a negligible artifact at the extremes, due to the omission of the first and the last onsets in the metronomes' timeseries. This artifact does not affect participants' behavior. The blue timeseries represent visually coupled conditions (1 and 2), where each participant tapped while seeing their partner's hand in continuous and discrete motion, respectively. Both exhibit parabolic modulation as a function of the distance from the in-phase pattern at the extremes of the metronomes' cycle, reaching a global minimum after the anti-phase midpoint. The red timeseries represent auditorily coupled conditions (3 and 4), where each participant tapped while hearing the continuous and discrete sonification of their partner's finger motion, respectively. Neither exhibited significant modulation as a function of the metronomes' cycles, but both were, on average, significantly higher than their visual counterparts.

differentiated based on Kinematics. Whereas in both Auditory conditions they fluctuate around a higher mean level, in Visual conditions they form a valley with global minimum after the anti-phase midpoint. Although visual inspection may suggest that the modulation is less prominent in condition of visual occlusion, the 3-way interaction between Modality, Kinematics and either polynomial terms does not reach statistical significance (linear: Estimate = 238.964, SE = 480.515, p = 0.619; quadratic: Estimate = 615.946, SE = 665.241, p = 0.354).

Table 1 presents the parameter estimates from the fixed effects model, their standard errors, and the associated p-values, with reference to the recurrence score. In this analytical framework, parameter estimates offer a measure of effect size with straightforward interpretation for linear and non-linear changes over time. By examining the interaction effect of Modality and Kinematics on the polynomial terms of Time, we could quantify the specific impact of access to movement kinematics on the evolution of the recurrence score across sensory modalities.

## 3. Discussion

In the present work, we investigated the influence of kinematic information on the dynamics of interpersonal synchronization, when the information is processed in visual and auditory modalities. Our aim was to explain previous findings demonstrating that attractor dynamics depend on the sensory modality mediating informational coupling [9]. We hypothesized that the differences observed in our previous study could be explained by the preferential access of visual modality to continuous kinematic information conveyed by the partner's movement. Specifically, we predicted that differences in coordination patterns would cancel out, provided a reliable cross-modal match in the information available across modalities. In order to test this hypothesis, we balanced the access to kinematics across modalities by augmenting information in the auditory domain via sonification and reducing it in the visual domain via occlusion. In doing so, we disentangled the contribution of the specific perceptual system processing information from the physical properties of the information itself.

Contrary to our hypothesis, the results supported a genuine modality-dependency of attractor dynamics. Specifically, we found that the attractor landscape remained invariant to the kinematic properties of the partner's movement. As shown in Fig. 2, the same pattern reported in conditions of visual coupling in Refs. [9,11] persisted even when occlusion impeded continuous access to kinematics. Conversely, the continuous sonification of movement did not modulate the interaction over time, resulting in high recurrence scores throughout the entire drifting metronomes' cycle. In summary, visually mediated interactions tended to result in recurrent transitions between cooperation and competition processes modulated by opposing attractors [28], whereas auditorily mediated interactions tended to result in constant cooperation throughout the task. We initially predicted that continuous sonification would enable participants to discern when their partner's effector was sufficiently distant from in-phase locking, allowing them to exploit the anti-phase region to decouple and pursue independent trajectories. Notably, this advantage was not exploited in the auditory modality.

We must admit, it could be argued that the adopted sonification strategy was simply not effective in conveying a continuous representation of the fingertip position with respect to the target, and consequently encoding its kinematics over time. Whilst this is certainly a possibility, our choice of frequency modulation (FM) was carefully considered and well-founded. From the perspective of the listener, the connection with low-level physical features is essential to effectively decode movement features from sound [30]. FM was preferred as sonification strategy over amplitude modulation (AM) because it was proven effective in influencing auditory-motor coordination [17]. Whilst there seems to be no or very little difference across AM and FM in terms of stabilizing auditory-motor coordination [21], FM resulted in better precision and rapidity when adopted in a 1-D guidance task [31], and was proven to effectively mediate kinematic information to a perceiver when explicitly used to substitute visual information [24]. Finally, pitch induces the best cross-modal correspondence with the visual dimensions of position and direction of movement along the vertical axis [1,32,33], which was the critical spatial dimension in our finger-tapping task. The mapping between the distance from the pad and the pitch was inverted, so that an ascending pitch was perceived as the fingertip approached the target. In continuously frequency modulated sounds, the detection of peaks is systematically better as compared to the troughs [34–37], which makes an ascending modulation

**Table 1**
Joint recurrence score. Orthogonal polynomials model summary.

| Predictors | Recurrence score (N = 19) | | |
|---|---|---|---|
| | *Estimate* | *SE* | *p* |
| (Intercept) | 2387.152 | 168.387 | 0.000 |
| Time | −155.326 | 265.929 | 0.559 |
| Time$^2$ | 217.347 | 398.003 | 0.585 |
| Modality | −532.994* | 207.089 | 0.010 |
| Kinematics | −75.791 | 207.089 | 0.714 |
| Time:Modality | −1049.894** | 339.775 | 0.002 |
| Time$^2$:Modality | 1230.757** | 470.396 | 0.008 |
| Time:Kinematics | −63.972 | 339.775 | 0.851 |
| Time$^2$:Kinematics | 138.860 | 470.396 | 0.767 |
| Modality: Kinematics | 257.607 | 292.868 | 0.379 |
| Time: Modality: Kinematics | 238.964 | 480.515 | 0.619 |
| Time$^2$: Modality: Kinematics | 615.946 | 665.241 | 0.354 |
| | *\*p < 0.05* | *\*\*p < 0.01* | *\*\*\*p < 0.001* |

more encoding motion leading to a discrete event. This choice was supported by extensive piloting, and is in line with the natural correspondence with Doppler effect, ecologically experienced as sound sources approach a target. We do not exclude that results might have changed if a more effective modulation had been implemented. It should also be noted that, in alternative to filling in the mute phases between the discrete taps, one could have opted for removing the 'tick' sound at the time of impact in the condition of continuous auditorily coupling. While we are aware that sonifying the impact along with the FM could have triggered neural circuits for discrete timing, we consider our choice to be the best auditory counterpart to the visual conditions, respectively adding and removing continuous kinematics leading to the event of the impact.

Having elucidated the rationale behind our sonification strategy, it is evident that the domain of auditory coupling through sonification offers a wide range of possibilities, whose potential impact on interpersonal coordination merits further inquiry. In contrast, visual occlusion can only be implemented by physically obstructing the access to kinematics. In absence of obvious alternatives to our approach, we found it particularly surprising that visually occluding kinematics did not significantly disrupt the attractor pattern, nor resulted in overall weaker coupling among partners as compared to the continuous condition. Previous research has demonstrated that exposure to the motion of periodic stimuli significantly improves visuomotor synchronization by engaging error correction mechanisms [38,39], and enriching the stimulus with a variable velocity profile further reduces variability [40,41] up to matching the performance with auditory stimuli presented at moderate tempi [42]. The invariance between continuous and discrete conditions led us to conclude that modality-specific differences may stem from the inherent characteristics of the two perceptual systems and from their interactions with timing and motor mechanisms, particularly in presence of concomitant streams of information.

Irrespective of the interpersonal domain, the primary source of divergence in synchronization performance is typically ascribed to the distinct specializations of visual and auditory modalities, which are respectively associated with spatial and temporal processing [26]. These two forms of processing have their neural bases in distinct native networks which include their preferred modality-biased areas [43–45]. The specialization of each network becomes evident when the task imposes one form of processing (spatial or temporal) in the non-preferred sensory modality (visual or auditory, respectively): in this kind of scenario, cross-modal recruitment can take place to compensate the inefficiency of the non-native network [43]. This implies that in presence of multimodal interference during temporal processing, task demands are greater for the visual system, leading the auditory network to take over the processing. This functional asymmetry can explain why individual behaviors are temporally attracted more strongly by auditorily perceived actions, irrespective of the attractor landscape.

The auditory dominance in timing is thought to stem from its stronger coupling to the motor system, which is evident in greater activation in motor structures such as the SMA and premotor cortex [46], as well as increased activation in the putamen [47] independent of stimulus motion [48]. Furthermore, auditory rhythms provide a selective benefit in pacing motor behavior when compensating for basal ganglia impairment in Parkinson's Disease [49,50]. It should be mentioned that in spite of the abundant evidence supporting anatomo-functional connectivity as a solid ground for temporal specialization, the dynamics of neural oscillations do not appear to be strictly modality-specific. While activity in the beta band has been traditionally linked to timing prediction in the auditory domain [51–54], recent research has also associated beta activity with timing predictions within the visual system in response to visual rhythms [26], as well as in cortico-kinematic coupling during real and imagined sensorimotor synchronization [55]. In a dyadic setting, we observed modulation of beta power as a function of the partner's movement cycles during both visual and auditory couplings [10], extending the notion of supramodality for this neural dynamic to the interpersonal domain.

Turning our attention to the temporal competences of the visual system, it is notable that despite its spatial dominance [26], visual cortex exhibits activation specific to interval timing [56], duration perception [57] and prediction of rhythmic onsets [58]. However, these capabilities face increased computational demand due to the inherent complexity of spatial processing within visual scenes, irrespective of stimulus continuity. Given this complexity, coupled with a weaker connection to the motor system, visual targets become less likely to trigger error correction mechanisms and drive adaptive behavior [26]. Considering these differences, when visual and auditory sequences presented out of phase directly compete in instructed synchronization, the auditory sequence takes precedence over the visual one, rather than the other way around [59,60]. Such multimodal competition is directly reflected in our dyadic paradigm, where whilst sounds produced by the partner entrain individual behaviors at all times, the weaker visual coupling loses traction as the drifting metronomes drive partners away from the in-phase attractor. The conclusion we take is that the cooperation process driving dyadic entrainment, when mediated by visual coupling, strongly relies on attractor dynamics to compensate for the timing disadvantage in this modality.

How do our findings align with major theories of interpersonal synchronization [22]? Predictive accounts propose that rhythmic behavior is driven by the continuous updating of internal predictive models of the other, facilitating coupled individuals to converge towards a synchronized state, with the aim of minimizing energy via error correction [23]. Within this framework, our 'kinematic hypothesis' anticipated that continuous and discrete access to informational observables from a partner's movement would lead to different updates of predictive models. With continuous access, the model can be updated continuously at the sampling rate of the specific perceptual system, thereby providing a fine-grained representation of the movement cycle. In contrast, when the access is discrete, the estimation of the next tap is based on prior knowledge of the tapping rate, resulting in irregular updates dependent on the actual occurrence of the next event. Contrary to a Bayesian account premised on supramodal updating [61], our findings suggest that spatial and temporal information are differentially weighted according to the sensory input channel. For example, the 'ventriloquist effect' demonstrates the dominance of visual information in spatial processing [62–64], while the 'flash-sound illusion' highlights auditory supremacy in processing temporal structure [65–68]. Continuous availability of informational observables might be more heavily weighted when processed visually, as they convey information on the spatial positioning of the partner's effector. The spatial processing of continuous sonification might be minimal, therefore having a negligible impact on behavior.

From a different perspective, within a dynamical systems framework [69–71], our 'kinematic hypothesis' predicted that continuous information would override the action of internal time-keepers [72], driving behavior through simple sensorimotor interactions across modalities. This form of emergent timing would allow participants to rely on the inherent properties of the partner's movement and of the feedback itself, without employing cognitive resources to compute intervals based on discrete events [25,73–76]. However, we should note that the perception of oscillating objects (such as the partner's hand) varies across sensory pathways. The spatial layout of these objects is mapped onto the retina and preserved throughout the visual pathway [77,78], whereas the receptor layout in the inner ear encodes the position of the target by combining localization and spectral cues [79,80]. These differences reasonably lead to varied information transfer from perceptual to motor oscillators [71], suggesting that the effector is driven by unique dynamics depending on the sensory input. Furthermore, the contribution of active sampling strategies unique to each sensory system cannot be neglected, particularly eye-tracking behavior, which has been shown to strengthen visuomotor coordination and modulate the effects of kinematics [81,82]. One study even suggests that interpersonal motor contagion can be suppressed when controlling for active tracking [83]. While we did not record eye movements in our study, it is unlikely that finger-tapping behavior elicited significant tracking due to the small amplitude of movements. Nonetheless, caution is warranted when generalizing our findings to other behaviors, especially those involving large-amplitude movements that may engage eye tracking to a larger extent. In such scenarios, recording eye movements and considering their effects on visuomotor coordination becomes essential.

In conclusion, our findings imply that the intrinsic dynamics of sensory perception and their interaction with the motor system play a pivotal role in interpersonal synchronization. The present study enriches the existing body of literature on interpersonal coordination by emphasizing the nature and accessibility of kinematic information, comparing the specific contributions of visual and auditory couplings. A thorough understanding of these facets is critical for designing and optimizing motor training and rehabilitation protocols that involve interaction between an individual and a trainer or another partner. While sonification has proven its worth for bio-feedback applications centered on the individual, both in sports [18,84–88] and rehabilitation [89–92], its potential as a mediator and enhancer of dyadic coupling remains relatively unexplored. We urge further investigation into this area to broaden our understanding of the various dimensions of auditory coupling. Grasping its potential and limitations in supporting motor coordination is essential for evaluating its validity, as well as for complementing visual input during dyadic interactions.

## 4. Methods

*Participants.* A total of 40 right-handed participants took part in the study, including 20 females and 20 males. The mean age was 31.2 years, with a standard deviation of 6.8 years. Participants were initially divided based on gender and subsequently randomly paired into 20 dyads. Upon recruitment, all participants reported not being professional musicians and having no history of neurological, major medical, or psychiatric disorders. With the exception of one dyad, participants indicated they were not familiar with their assigned partner.

*Experimental Apparatus.* Participants were seated across from each other at the same table, on chairs equipped with armrests. By resting their elbows on the armrests, participants eliminated any potential tactile or proprioceptive coupling between them due to table vibrations. Each participant was assigned a pad and instructed to tap on it using their right index finger in synchronization with a metronome. Depending on the experimental condition, the metronome provided either an auditory cue or a flickering light. Two wooden screens were placed perpendicularly on the table to manipulate visual coupling between the partners. The frontal plane was equipped with adjustable sliding windows to regulate visual access to the partner's hand movements based on height and position of the participant. Each participant had to wear a glove with an infrared reflective marker on the finger-tip, used to track the finger's position during the task. A Qualisys Motion Capture system, consisting of 8 Miqus M3 cameras, 1 Miqus RGB camera and the 'Qualisys Track Manager 2020' software was used to track the position of the hand and tapping pads. A Focusrite RedNet 2 Dante Audio Interface was used to deliver independent audio channels to each participant through in-ear plugs, with Ableton Live 10 serving as the main interface for stimuli presentation, sonification of finger-taps, and real-time routing. The same MIDI tracks controlled the metronomes across conditions by triggering either an audio sample or a flickering LED. A Teensy 3.2 microcontroller functioned as a serial/MIDI hub in the setup, detecting tapping onsets with <1 ms resolution based on analog input from strain gauge pressure sensors installed within the pads. Every time a metronome onset was presented to a participant, a MIDI message was sent to the Teensy device to log the metronome's timeseries and control the voltage of the LEDs when necessary. Participants were monitored from a control desk placed behind a curtain using the Miqus RGB camera to ensure compliance with instructions.

*Experimental design.* The study was designed in a Modality (Visual, Auditory) x Kinematics (Continuous, Discrete) $2 \times 2$ factorial structure, as represented in Fig. 1. In all conditions, participants were always exposed to the partner's movements in one modality while they were paced by a metronome in the other modality. They were instructed to neglect the partner and focus on following their own metronome at all times. The design resulted in the following experimental conditions. *1. Visually Coupled – Continuous*: participants tapped along with an auditory metronome, while looking at the partner's hand tapping. The hand of the partner was fully visible, while the view of their own hand was hidden by a component of the screen. *2. Visually Coupled – Discrete*: participants tapped along with an auditory metronome, while looking at the partner's hand tapping. The hand of the partner was hidden by a sliding window, adjusted so that only the fingertip was visible at the time of impact. The view of their own hand was hidden by the other component of the screen. *3. Auditorily Coupled – Continuous*: participants tapped along with a flickering LED, while hearing the sonification of the partner's tapping. The sonification consisted of a continuous tone, modulated in frequency by the distance between the fingertip and the pad. A discrete 'tick' sound was triggered at the time of impact. The view of their own hand was hidden by a component of the screen. *4. Auditorily Coupled – Discrete*: participants tapped along with a flickering LED, while hearing the sonification of the partner's tapping. The sonification consisted of a discrete 'tick' sound triggered at the time of impact. The view of their own hand was hidden by

a component of the screen.

*Experimental procedure.* Before participants arrived in the lab, the experimenters calibrated the motion capture system. Upon arrival in the lab, participants read all information related to the study and signed the informed consent form. Demographic data were collected with a questionnaire, and the Edinburgh inventory [93] was administered to assess participants' right-handedness. The Interpersonal Reactivity Index (IRI) [94] was administered as a self-report measure of empathy and its subscales. Participants' preferred spontaneous tempo was calculated through 30 s of self-paced finger-tapping on a dedicated smartphone app (www.beatsperminuteonline.com). After wearing the glove for tracking the finger's position, participants wore the earbuds. Volume levels were adjusted for each participant before the experiment, and pink noise was regulated to suppress any environmental sound. After instructing the participants on how to perform finger-taps in synchronization with the metronome, they underwent a 30-s familiarization session. Subsequently, participants were asked to place their finger on the pad and then hold it at the highest point of their comfortable range of motion, while the experimenter recorded the positions on the software to normalize the motion range for continuous sonification. In visually coupled conditions, the lights of the room were lit at the minimum level and two diffused directional lights were oriented parallel to the table, to prevent the lights coming from the ceiling to cast the shadow of the moving hand on the pad. In condition of visual occlusion, the eight of the sliding windows on the wooden screen was adjusted based on the height and position of the participants, so that only the fingertip of the partner was visible at the moment of impact with the pad. The importance of maintaining the position and the style of tapping was stressed in the instructions. All dyads confirmed during debriefing that the manipulation was successful. In either sensory modality (ticking auditory metronome or flickering visual metronome), each participant was cued with a different tempo (1.67 Hz and 1.64 Hz), such that the relative phase between the metronomes began at $0°$, increasing regularly in $5.6°$ increments [9]. A full cycle took 39 s to complete, with 10 consecutive cycles performed in each experimental condition. Participants were clearly instructed to ignore their partner at all times, and tap along with their assigned metronome. During breaks between experimental conditions, participants provided subjective self-reports on various aspects of the task by expressing agreement on a scale from 1 ("Completely disagree") to 7 ("Completely agree") using a custom-made battery of 12 Likert items. Data collected from the questionnaires are not presented in this paper.

*Motion capture and sonification.* The real-time sonification of the finger-tapping motion was implemented as follows. Five reflective markers for infrared light were attached to the pad of each participant in different configurations, such that 2 rigid bodies were identified in the motion capture recording. One extra marker was placed in the center of the pads for the creation of rigid bodies, defined as virtual marker, and physically removed to define the origin of the XYZ space. The directions of the axes were relative to the participants' bodies (X pointing outwards, Y pointing to the front, Z pointing to the upside). This way, two XY planes were defined, each lying on the surface of the respective pad. Each participant performed the experiment wearing a glove equipped with a marker on the back of the middle phalanx of the index finger. During finger-tapping, we sonified the Euclidean distance of the marker from the XY plane of the assigned pad. This parameter was mapped onto the frequency of a sinusoidal waveform using the software Ableton Live, resulting in a continuous frequency modulation (FM) of the audio signal. Specifically, a downward movement towards the pad increased the pitch, while an upward movement away from the pad decreased it. The modulation ranged from the note C4 (261 Hz) to the note C5 (523 Hz), with the pitch range normalized based on each participant's range of motion during the calibration phase. This sonification was implemented exclusively in Condition 3 (Auditory Coupling, Continuous) to bridge the temporal gaps between tapping onsets. The moment of impact was sonified by triggering the audio sample of a discrete metronome sound within Ableton Live for both Condition 3 (Auditory Coupling, Continuous) and Condition 4 (Auditory Coupling, Discrete).

## 4.1. Data analysis

For the most part, the pipeline hereby described is a translation of the methods described in Refs. [9,10], applied to the current experimental design. The content of this paragraph is adapted from the original work with the authors' consent. However, the values for the embedding dimension and delay differ from the original study as they were optimized for the current dataset.

*Pre-processing.* Throughout the duration of 10 consecutive metronome cycles, each dyad's partners were expected to produce 650 and 640 tapping onsets, respectively, over a total period of 390 s. Onsets occurring less than 350 ms from the previous one were considered false positives and subsequently removed, as participants could occasionally press the pad for an extended duration or inadvertently rest their hand on it. From the entire sample, 89 false positives were eliminated, accounting for 0.086 % of all data points. The refined timeseries were then interpolated using a sine function at a 1 kHz sampling rate, yielding an estimate of the oscillators' positions within the cycle at a temporal resolution of 1 ms. The tap prior to the first metronome onset and the final tap following the last metronome onset were incorporated into the interpolation. Subsequently, data points outside the metronomes' timeseries boundaries were discarded. This procedure ensured equally sized timeseries without data loss, a prerequisite for the implementation of joint recurrence quantification analysis (JRQA; see the following paragraph). The interpolation choice is conceptually supported by the modeling of coupled oscillator systems in the context of finger-tapping studies [9,10,71,95]. Lastly, timeseries were down-sampled by a factor of 4 to render the computation of recurrence plots (RPs) more manageable. As demonstrated in our previous report, JRQA results are not sensitive to the choice of down-sampling factor. Dyad #15 was excluded from the analysis, because the Edinburgh Inventory revealed that one participant was effectively left-handed. Furthermore, the pad of one participant did not work for the whole duration of Condition #1, resulting in the loss of all data collected in that condition.

*Phase-space reconstruction.* Following Takens' embedding theorem [96], we reconstructed the phase space of individual finger-tapping behaviors. This involved utilizing time-delayed copies of the input timeseries $u_k$, while implementing an embedding dimension $m$ and a time delay $\tau$.

$$\vec{x}(t) = \vec{x}_i = (u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}), t = i\Delta t \tag{1}$$

where $x{\rightarrow}(t)$ represents the vector of reconstructed states within the phase-space at time $t$. Optimal parameters for the time-delayed embedding were calculated for each participant, for the duration of each individual metronome cycle across all experimental conditions. The resulting average value of the parameters was applied to all individual timeseries. The reason is that to compare recurrence rates across conditions at a group level, the embedding procedure must remain consistent among participants (e.g., see Refs. [9, 97], for examples of parameter selections in a factorial design). Firstly, we chose the delay $\tau$ as the first local minimum of the mutual information index [98] in relation to the delay. This method minimized the timeseries self-similarity, extracting almost orthogonal components and preventing the attractor from folding onto itself [99]. The mean delay value was determined to be $\tau = 7$. Next, we identified the number of embedding dimensions using the false nearest neighbor method [100]. Specifically, we incrementally unfolded the timeseries into higher dimensions until data points no longer overlapped spuriously, identifying an optimal mean embedding of $m = 2$. Lastly, in line with existing literature, the maximum threshold for considering two adjacent points as recurrent was set at 10 % of the maximum phase-space diameter [27].

**Joint recurrence plots (JRPs).** A recurrence plot, denoted as $R_{i,j}$, is a square array employed to represent and quantify the recurrence of states within a system's phase space [101]. For each point in the phase space trajectory:

$$\vec{x}_i \ (i = 1, \ldots, N; \ N = n - (m - 1) \tau) \tag{2}$$

the proximity to another trajectory point, $x{\rightarrow}_j$, was assessed based on a neighborhood threshold. Individual recurrence plots are computed using the following equation:

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|x_i - x_j\|) \tag{3}$$

Here, $\varepsilon$ represents the neighborhood threshold, $\| \cdot \|$ is the Euclidean norm (indicating the distance between two vectors), and $\Theta$ is the Heaviside step function. Each phase-space yields a square matrix with 1s where the distance $\| \cdot \|$ is less than the threshold $\varepsilon$, and 0s for other elements.

A joint recurrence plot (JRP) was computed for each dyad by overlapping individual RPs pairwise, retaining 1s only where both plots exhibited recurrence. Each JRP was essentially the Hadamard product of the first and second participants' recurrence plots. The *CRP Toolbox* for Matlab® was used to compute JRPs [102]. By aggregating the JRPs of the 10 trials (i.e., metronome cycles) for each experimental condition, a 2-D matrix was obtained, with each entry representing the recurrence count in the corresponding cycle region across all trials. A 1-D vector recurrence score was derived by summing the row counts for each matrix column. This vector signifies a density measure of coupled behavior instances throughout the metronomes' cycle. The recurrence score scale relies on the JRP size, which, in turn, depends on the embedding procedure. This necessitates setting consistent parameters across the entire sample. To enhance the signal-to-noise ratio and prevent oversampling, the resulting timeseries were reduced to 64 bins by averaging recurrence scores for equally sized, consecutive time periods. The interval size was determined by the slower metronome's increments, which provided a regular subdivision inherent to the experimental trials.

All processing steps were conducted in Matlab®. Our approach was chosen over the moving window-based JRQA version to avoid low-pass filtering effects on the timeseries, which could impede result interpretation. Specifically, a moving window can introduce phase distortion in the timeseries based on window size, making it unreliable for detecting attractor points across the landscape.

**Statistical models.** The recurrence score served as the response variable in a mixed-effects model, which included Coupling and Perspective as factors and Time as a continuous predictor, expressed through the indexes of the metronome steps (from 1 to 64). Due to the non-linear time course observed in coupled conditions, we employed the method of orthogonal polynomials [9,11,29,103], incorporating linear and quadratic functions of Time into our model. Dyads and interactions between Dyads and the factors were modeled as random effects on all polynomial terms to accommodate individual variability in synchronization abilities and individual susceptibility to coupling across experimental manipulations. The random effects structure was adopted to minimize false alarm rates without significant power loss [104]. Guided by our prior study and the examination of empirical curves from the current dataset, we confined the polynomial model to the 2nd order as the most parsimonious solution. Within this analytical framework, the intercept is considered a 'zero-order' polynomial, as it displays zero changes in any direction. Significant changes in direction signify modulation by the task's temporal structure, enabling us to quantify the influence of attractor points as the dyad diverged from the horizontal trajectory while transitioning over the expected critical regions.

The full model's formula is as follows:

$$Recurrence \sim (Time + Time^2) * Modality * Kinematics + (Time + Time^2 \mid Dyad) + (Time + Time^2 \mid Dyad : Modality : Kinematics)$$

Statistical analyses were carried out in R (version 4.0.3). The *lme4* package [105] was used for model fitting.

## Ethics and consent

The experiment received approval from the Ethics Committee of Ghent University on November 15th, 2022 (Faculty of Arts and Philosophy; ethical approval number FLW_2022-34), and informed written consent was obtained from each participant before commencing the experiment. Participants were compensated with a 20€ coupon for their time.

## Data availability

The datasets and the scripts used to perform the analyses are publicly available at the following Mendeley repository:https://data.mendeley.com/preview/y4kt5924fz?a=61cc7d02-b3b6-441b-86d6-2e19f61d0473.

## CRediT authorship contribution statement

**Mattia Rosso:** Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Canan Nuran Gener:** Writing – review & editing, Investigation. **Bart Moens:** Writing – review & editing, Software, Resources. **Pieter-Jan Maes:** Writing – review & editing, Supervision. **Marc Leman:** Writing – review & editing, Supervision, Project administration.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

## References

[1]  C. Spence, Crossmodal correspondences: a tutorial review, Atten. Percept. Psychophys. 73 (2011) 971–995.
[2]  J. Phillips-Silver, C.A. Aktipis, G.A. Bryant, The ecology of entrainment: Foundations of coordinated rhythmic movement, Music Percept. 28 (2010) 3–14.
[3]  A.O. Effenberg, Movement sonification: effects on perception and action, IEEE Multimedia 12 (2005) 53–59.
[4]  R.C. Schmidt, C. Carello, M.T. Turvey, Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people, J. Exp. Psychol. Hum. Percept. Perform. 16 (1990) 227–247.
[5]  P.N. Kugler, M.T. Turvey, Information, Natural Law, and the Self-Assembly of Rhythmic Movement, Routledge, 2015.
[6]  E. Tognoli, M. Zhang, A. Fuchs, C. Beetle, J.A.S. Kelso, Coordination dynamics: a foundation for understanding social behavior, Front. Hum. Neurosci. 14 (2020) 317.
[7]  G. Knoblich, N. Sebanz, Evolving intentions for social interaction: from entrainment to joint action, Philos. Trans. R. Soc. Lond. B Biol. Sci. 363 (2008) 2021–2031.
[8]  J.J. Gibson, The Ecological Approach to Visual Perception, Classic Edition, Psychology Press, 2014.
[9]  M. Rosso, P.J. Maes, M. Leman, Modality-specific attractor dynamics in dyadic entrainment, Sci. Rep. 11 (2021) 18355.
[10]  M. Rosso, O.A. Heggli, P.J. Maes, P. Vuust, M. Leman, Mutual beta power modulation in dyadic entrainment, Neuroimage 257 (2022) 119326.
[11]  M. Rosso, B. Van Kerrebroeck, P.J. Maes, M. Leman, Embodied perspective-taking enhances interpersonal synchronization: A body-swap study, Iscience 26 (11) (2023).
[12]  J. Yamanishi, M. Kawato, R. Suzuki, Two coupled oscillators as a model for the coordinated finger tapping by both hands, Biol. Cybern. 37 (1980) 219–225.
[13]  B. Tuller, J.A. Kelso, Environmentally-specified patterns of movement coordination in normal and split-brain subjects, Exp. Brain Res. 75 (1989) 306–316.
[14]  L.E. Marks, Cross-modal interactions in speeded classification, in: G.A. Calvert (Ed.), The Handbook of Multisensory Processes, 2004, pp. 85–105 (Boston Review, xvii).
[15]  A. Gallace, C. Spence, Multisensory synesthetic interactions in the speeded classification of visual size, Percept. Psychophys. 68 (2006) 1191–1203.
[16]  C.V. Parise, C. Spence, "When birds of a feather flock together": synesthetic correspondences modulate audiovisual integration in non-synesthetes, PLoS One 4 (2009) e5664.
[17]  M. Varlet, L. Marin, J. Issartel, R.C. Schmidt, B.G. Bardy, Continuity of visual and auditory rhythms influences sensorimotor coordination, PLoS One 7 (2012) e44082.
[18]  A.O. Effenberg, U. Fehse, G. Schmitz, B. Krueger, H. Mechling, Movement sonification: effects on motor learning beyond rhythmic adjustments, Front. Neurosci. 10 (2016) 219.
[19]  F. Bevilacqua, E.O. Boyer, J. Françoise, O. Houix, P. Susini, A. Roby-Brami, S. Hanneton, Sensori-motor learning with movement sonification: perspectives from recent interdisciplinary studies, Front. Neurosci. 10 (2016) 385.
[20]  K.R. Agres, R.S. Schaefer, A. Volk, S. van Hooren, A. Holzapfel, S. Dalla Bella, M. Müller, M. de Witte, D. Herremans, R. Ramirez Melendez, et al., Music, computing, and health: a roadmap for the current and future roles of music technology for health care and well-being, Music & Science 4 (2021) 2059204321997709.
[21]  M.W.M. Rodger, C.M. Craig, Timing movements to interval durations specified by discrete or continuous sounds, Exp. Brain Res. 214 (2011) 393–402.
[22]  C. Palmer, A.P. Demos, Are we in time? How predictive coding and dynamical systems explain musical synchrony, Curr. Dir. Psychol. Sci. 31 (2022) 147–153.
[23]  L. Koban, A. Ramamoorthy, I. Konvalinka, Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle, Soc. Neurosci. 14 (2019) 1–9.
[24]  A.O. Effenberg, G. Schmitz, Acceleration and deceleration at constant speed: systematic modulation of motion perception by kinematic sonification, Ann. N. Y. Acad. Sci. 1425 (2018) 52–69.
[25]  P.-J. Maes, M. Giacofci, M. Leman, Auditory and motor contributions to the timing of melodies under cognitive load, J. Exp. Psychol. Hum. Percept. Perform. 41 (2015) 1336–1352.
[26]  D.C. Comstock, M.J. Hove, R. Balasubramaniam, Sensorimotor synchronization with auditory and visual modalities: behavioral and neural differences, Front. Comput. Neurosci. 12 (2018) 53.
[27]  N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. 438 (2007) 237–329.
[28]  K.L. Marsh, M.J. Richardson, R.C. Schmidt, Social connection through joint action and interpersonal coordination, Top. Cogn. Sci. 1 (2009) 320–339.
[29]  D. Mirman, Growth Curve Analysis and Visualization Using R, CRC Press), 2017.
[30]  R. Bresin, M. Mancini, L. Elblaus, E. Frid, Sonification of the self vs. sonification of the other: differences in the sonification of performed vs. observed simple hand movements, Int. J. Hum. Comput. Stud. 144 (2020) 102500.

[31] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, R. Kronland-Martinet, Comparison and evaluation of sonification strategies for guidance tasks, IEEE Trans. Multimed. 18 (2016) 674–686.

[32] K.K. Evans, A. Treisman, Natural cross-modal mappings between visual and auditory features, J. Vis. 10 (6) (2010) 1–12.

[33] F. Maeda, R. Kanai, S. Shimojo, Changing pitch induced visual motion illusion, Curr. Biol. 14 (2004) R990–R991.

[34] L. Demany, K.I. McAnally, The perception of frequency peaks and troughs in wide frequency modulations, J. Acoust. Soc. Am. 96 (1994) 706–715.

[35] L. Demany, S. Clément, The perception of frequency peaks and troughs in wide frequency modulations. II. Effects of frequency register, stimulus uncertainty, and intensity, J. Acoust. Soc. Am. 97 (1995) 2454–2459.

[36] L. Demany, S. Clément, The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform, J. Acoust. Soc. Am. 102 (1997) 2935–2944.

[37] A. de Cheveigné, A model of the perceptual asymmetry between peaks and troughs of frequency modulation, J. Acoust. Soc. Am. 107 (2000) 2645–2656.

[38] M.J. Hove, M.J. Spivey, C.L. Krumhansl, Compatibility of motion facilitates visuomotor synchronization, J. Exp. Psychol. Hum. Percept. Perform. 36 (2010) 1525–1534.

[39] M.J. Hove, P.E. Keller, Spatiotemporal relations and movement trajectories in visuomotor synchronization, Music Percept. 28 (2010) 15–26.

[40] M.J. Hove, J.R. Iversen, A. Zhang, B.H. Repp, Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome, Psychol. Res. 77 (2013) 388–398.

[41] J.R. Iversen, A.D. Patel, B. Nicodemus, K. Emmorey, Synchronization to auditory and visual rhythms in hearing and deaf individuals, Cognition 134 (2015) 232–244.

[42] L. Gan, Y. Huang, L. Zhou, C. Qian, X. Wu, Synchronization to a bouncing ball with a realistic motion trajectory, Sci. Rep. 5 (2015) 11974.

[43] S.W. Michalka, L. Kong, M.L. Rosen, B.G. Shinn-Cunningham, D.C. Somers, Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks, Neuron 87 (2015) 882–892.

[44] A.L. Noyce, N. Cestero, S.W. Michalka, B.G. Shinn-Cunningham, D.C. Somers, Sensory-biased and multiple-demand processing in human lateral frontal cortex, J. Neurosci. 37 (2017) 8755–8766.

[45] A.L. Noyce, R.W. Lefco, J.A. Brissenden, S.M. Tobyne, B.G. Shinn-Cunningham, D.C. Somers, Extended frontal networks for visual and auditory working memory, Cerebr. Cortex 32 (2022) 855–869.

[46] L. Jäncke, R. Loose, K. Lutz, K. Specht, N.J. Shah, Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli, Brain Res. Cogn. Brain Res. 10 (2000) 51–66.

[47] J.A. Grahn, M.J. Henry, J.D. McAuley, FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa, Neuroimage 54 (2011) 1231–1243.

[48] M.J. Hove, M.T. Fairhurst, S.A. Kotz, P.E. Keller, Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness, Neuroimage 67 (2013) 313–321.

[49] M.J. de Dreu, A.S.D. van der Wilk, E. Poppe, G. Kwakkel, E.E.H. van Wegen, Rehabilitation, exercise therapy and music in patients with Parkinson's disease: a meta-analysis of the effects of music-based movement therapy on walking ability, balance and quality of life, Parkinsonism Relat. Disorders 18 (2012) S114–S119.

[50] A. Ashoori, D.M. Eagleman, J. Jankovic, Effects of auditory rhythm and music on gait disturbances in Parkinson's disease, Front. Neurol. 6 (2015) 234.

[51] T. Fujioka, L.J. Trainor, E.W. Large, B. Ross, Beta and gamma rhythms in human auditory cortex during musical beat processing, Ann. N. Y. Acad. Sci. 1169 (1) (2009) 89–92.

[52] T. Fujioka, B. Ross, L.J. Trainor, Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery, J. Neurosci. 35 (2015) 15187–15198.

[53] J.R. Iversen, B.H. Repp, A.D. Patel, Top-down control of rhythm perception modulates early auditory responses, Ann. N. Y. Acad. Sci. 1169 (2009) 58–73.

[54] T. Fujioka, L.J. Trainor, E.W. Large, B. Ross, Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations, J. Neurosci. 32 (2012) 1791–1802.

[55] P. Nijhuis, P.E. Keller, S. Nozaradan, M. Varlet, Dynamic modulation of cortico-muscular coupling during real and imagined sensorimotor synchronisation, Neuroimage 238 (2021) 118209.

[56] M.G.H. Shuler, Timing in the visual cortex and its investigation, Curr Opin Behav Sci 8 (2016) 73–77.

[57] B. Zhou, S. Yang, L. Mao, S. Han, Visual feature processing in the early visual cortex affects duration perception, J. Exp. Psychol. Gen. 143 (2014) 1893–1902.

[58] D.C. Comstock, R. Balasubramaniam, Neural responses to perturbations in visual and auditory metronomes during sensorimotor synchronization, Neuropsychologia 117 (2018) 55–66.

[59] B.H. Repp, A. Penel, Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences, J. Exp. Psychol. Hum. Percept. Perform. 28 (2002) 1085–1099.

[60] B.H. Repp, A. Penel, Rhythmic movement is attracted more strongly to auditory than to visual rhythms, Psychol. Res. 68 (2004) 252–270.

[61] S. Fait, S. Pighin, A. Passerini, F. Pavani, K. Tentori, Sensory and multisensory reasoning: is Bayesian updating modality-dependent? Cognition 234 (2023) 105355.

[62] M. Radeau, P. Bertelson, The after-effects of ventriloquism, Q. J. Exp. Psychol. 26 (1974) 63–71.

[63] J. Vroomen, P. Bertelson, B. de Gelder, Directing spatial attention towards the illusory location of a ventriloquized sound, Acta Psychol. 108 (2001) 21–33.

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

[65] L. Shams, Y. Kamitani, S. Shimojo, Illusions. What you see is what you hear, Nature 408 (2000) 788.

[66] D. Burr, M.S. Banks, M.C. Morrone, Auditory dominance over vision in the perception of interval duration, Exp. Brain Res. 198 (2009) 49–57.

[67] P. Bertelson, G. Aschersleben, Temporal ventriloquism: crossmodal interaction on the time dimension: 1. Evidence from auditory–visual temporal order judgment, Int. J. Psychophysiol. 50 (2003) 147–155.

[68] S. Morein-Zamir, S. Soto-Faraco, A. Kingstone, Auditory capture of vision: examining temporal ventriloquism, Brain Res. Cogn. Brain Res. 17 (2003) 154–163.

[69] N. Stepp, M.T. Turvey, On strong anticipation, Cognit. Syst. Res. 11 (2010) 148–164.

[70] A.P. Demos, H. Layeghi, M.M. Wanderley, C. Palmer, Staying together: a bidirectional delay-coupled approach to joint action, Cognit. Sci. 43 (2019) e12766.

[71] O.A. Heggli, J. Cabral, I. Konvalinka, P. Vuust, M.L. Kringelbach, A Kuramoto model of self-other integration across interpersonal synchronization strategies, PLoS Comput. Biol. 15 (2019) e1007422.

[72] R.B. Ivry, T.C. Richardson, Temporal control and coordination: the multiple timer model, Brain Cognit. 48 (2002) 117–132.

[73] E. Coorevits, P.-J. Maes, J. Six, M. Leman, The influence of performing gesture type on interpersonal musical timing, and the role of visual contact and tempo, Acta Psychol. 210 (2020) 103166.

[74] D. Delignières, L. Lemoine, K. Torre, Time intervals production in tapping and oscillatory motion, Hum. Mov. Sci. 23 (2004) 87–103.

[75] S.D. Robertson, H.N. Zelaznik, D.A. Lantero, K.G. Bojczyk, R.M. Spencer, J.G. Doffin, T. Schneidt, Correlations for timing consistency among tapping and drawing tasks: evidence against a single timing process for motor control, J. Exp. Psychol. Hum. Percept. Perform. 25 (1999) 1316–1330.

[76] K. Torre, R. Balasubramaniam, Two different processes for sensorimotor synchronization in continuous and discontinuous rhythmic movements, Exp. Brain Res. 199 (2009) 157–166.

[77] J.W. Triplett, M.T. Owens, J. Yamada, G. Lemke, J. Cang, M.P. Stryker, D.A. Feldheim, Retinal input instructs alignment of visual topographic maps, Cell 139 (2009) 175–185.

[78] L. Luo, J.G. Flanagan, Development of continuous and discrete neural maps, Neuron 56 (2007) 284–300.

[79] N.H. Salminen, H. Tiitinen, P.J.C. May, Auditory spatial processing in the human cortex, Neuroscientist 18 (2012) 602–612.

[80] J.C. Middlebrooks, D.M. Green, Sound localization by human listeners, Annu. Rev. Psychol. 42 (1991) 135–159.

[81] M.J. Richardson, K.L. Marsh, R.W. Isenhower, J.R.L. Goodman, R.C. Schmidt, Rocking together: dynamics of intentional and unintentional interpersonal coordination, Hum. Mov. Sci. 26 (2007) 867–891.

[82] R.C. Schmidt, M.J. Richardson, C. Arsenault, B. Galantucci, Visual tracking and entrainment to an environmental rhythm, J. Exp. Psychol. Hum. Percept. Perform. 33 (2007) 860–870.

[83] M.D. Constable, J. de Grosbois, T. Lung, L. Tremblay, J. Pratt, T.N. Welsh, Eye movements may cause motor contagion effects, Psychon. Bull. Rev. 24 (2017) 835–841.

[84] P.-J. Maes, V. Lorenzoni, J. Six, The SoundBike: musical sonification strategies to enhance cyclists' spontaneous synchronization to external music, J. Multimodal User Interfaces 13 (2019) 155–166.

[85] V. Lorenzoni, J. Staley, T. Marchant, K.E. Onderdijk, P.-J. Maes, M. Leman, The sonic instructor: a music-based biofeedback system for improving weightlifting technique, PLoS One 14 (2019) e0220915.

[86] V. Lorenzoni, P. Van den Berghe, P.-J. Maes, T. De Bie, D. De Clercq, M. Leman, Design and validation of an auditory biofeedback system for modification of running parameters, Journal on Multimodal User Interfaces 13 (2019) 167–180.

[87] V. Lorenzoni, P.-J. Maes, P. Van den Berghe, D. De Clercq, T. de Bie, M. Leman, A biofeedback music-sonification system for gait retraining, in: Proceedings of the 5th International Conference on Movement and Computing MOCO '18, Association for Computing Machinery), 2018, pp. 1–5.

[88] P. Van den Berghe, V. Lorenzoni, R. Derie, J. Six, J. Gerlo, M. Leman, D. De Clercq, Music-based biofeedback to reduce tibial shock in over-ground running: a proof-of-concept study, Sci. Rep. 11 (2021) 4091.

[89] L. Moumdjian, B. Moens, P.-J. Maes, J. Van Nieuwenhoven, B. Van Wijmeersch, M. Leman, P. Feys, Walking to music and metronome at various tempi in persons with multiple sclerosis: a basis for rehabilitation, Neurorehabil. Neural Repair 33 (2019) 464–475.

[90] P.-J. Maes, J. Buhmann, M. Leman, 3Mo: a model for music-based biofeedback, Front. Neurosci. 10 (2016) 548.

[91] B. Van Kerrebroeck, P.-J. Maes, A breathing sonification system to reduce stress during the COVID-19 pandemic, Front. Psychol. 12 (2021) 623110.

[92] F. Bevilacqua, I. Peyre, M. Segalen, P. Pradat-Diehl, V. Marchand-Pauvert, A. Roby-Brami, Exploring different movement sonification strategies for rehabilitation in clinical settings, in: Proceedings of the 5th International Conference on Movement and Computing MOCO '18, Association for Computing Machinery), 2018, pp. 1–6.

[93] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, Neuropsychologia 9 (1971) 97–113.

[94] M.H. Davis, Others, A Multidimensional Approach to Individual Differences in Empathy, 1980.

[95] M. Rosso, M. Leman, L. Moumdjian, Neural entrainment meets behavior: the stability index as a neural outcome measure of auditory-motor coupling, Front. Hum. Neurosci. 15 (2021) 668918.

[96] F. Takens, Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, Warwick 1980, Springer Berlin Heidelberg, 1981, pp. 366–381.

[97] O. Afsar, U. Tirnakli, N. Marwan, Recurrence Quantification Analysis at work: quasi-periodicity based interpretation of gait force profiles for patients with Parkinson disease, Sci. Rep. 8 (2018) 9102.

[98] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A Gen. Phys. 33 (1986) 1134–1140.

[99] E. Bradley, H. Kantz, Nonlinear time-series analysis revisited, Chaos 25 (2015) 097610.

[100] N.F. Rulkov, M.M. Sushchik, L.S. Tsimring, H.D. Abarbanel, Generalized synchronization of chaos in directionally coupled chaotic systems, Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 51 (1995) 980–994.

[101] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Others, Recurrence plots of dynamical systems, World Scientific Series on Nonlinear Science Series A 16 (1995) 441–446.

[102] N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots, Phys. Lett. A 302 (2002) 299–307.

[103] M. Rosso, B. Moens, M. Leman, L. Moumdjian, Neural entrainment underpins sensorimotor synchronization to dynamic rhythmic stimuli, Neuroimage 277 (2023) 120226.

[104] D.J. Barr, R. Levy, C. Scheepers, H.J. Tily, Random effects structure for confirmatory hypothesis testing: keep it maximal, J. Mem. Lang. 68 (2013) 255–278.

[105] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using Lme4. arXiv [stat.CO], 2014.