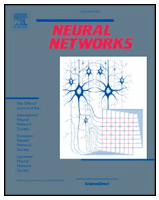
Neural Networks 201 (2026) 108960

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

BIVIP: a competitive neural network for bilateral visuo-haptic processing

Eleonore Federica Di Rosa [a,b,*], Aaisha Sheth [c], Jeffrey Min-In Yau [c], Laura Astolfi [b], Cristiano Cuppini [a]

[a] Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi," University of Bologna, Bologna, Italy
[b] Department of Computer Control and Management Engineering, Sapienza University of Rome, Rome, Italy
[c] Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Keywords:
Visuo-tactile
Switch cost
Cross-sensory competition
Bilateral touch
Neural network model

ABSTRACT

Visuo-haptic interactions are gaining interest in perception research because of their relevance to robotics and rehabilitation. While the benefits of integrating concurrent visuo-tactile cues are well-documented, the temporal dynamics of competitive interactions between vision and touch remain less understood. Building on extant empirical evidence of unilateral intersensory switching, we ask whether bilateral detection of visual and tactile stimuli can be explained by cross-sensory competition and explore the underlying sensorineural dynamics via a multi-level neural network model. Tested against results from a speeded reaction time task to unilateral and bilateral visual, tactile, and visuo-tactile targets, the network successfully predicts higher switch costs for sequential identification of unisensory stimuli of different modalities on the same hand rather than across hands. Our network also captures significant effects of inter-stimulus interval, trial type (i.e. repeat vs. switch), and modality type (i.e. visual, tactile, or visuo-tactile). Additionally, a sensitivity analysis on specific network connections differentiated between potential mechanisms underlying unisensory and multisensory processing of bilateral targets. Both network predictions and behavioral data reveal a slowdown in response to the repeated presentation of simultaneous visuo-tactile targets, suggesting that cross-sensory inhibition may contribute to offline influences of vision on bilateral tactile perception.

1. Introduction

Whether reaching for an object, navigating through an unfamiliar environment, or engaging in social interactions, humans rely on vision and touch to process incoming sensory signals from both sides of the body (Ehrsson, 2020). As perceiving congruent visuo-tactile stimulation enhances detection and reduces reaction times in healthy subjects (Holmes & Spence, 2005, Kennett et al., 2001, Spence et al., 2004, Rizzolatti et al., 1997), visuo-tactile interactions are increasingly exploited also for somatosensory feedback restoration following sensory-impairing conditions (Boesch et al., 2016; Dadarlat et al., 2015; Risso et al., 2019; Rognini et al., 2019). Beyond cooperating, however, vision and touch may engage in competitive interactions, with somatosensory perception becoming less accurate or even suppressed when visual distractors are presented on incongruent skin sites or spatial locations (Craig & Qian, 1997; Hartcher-O'Brien et al., 2008; Samad & Shams, 2016; Sarri et al., 2006; Soto-Faraco et al., 2004; Van Der Groen et al., 2013). As natural stimulation typically consists of temporally ordered sequences rather than single events, both cooperative and competitive visuo-tactile interactions unfold over time. The response to a target therefore reflects not only the congruency of the multisensory configuration, but also how the preceding stimulus modulates modality-specific excitability and sensorimotor readiness over the time interval between events.

Under this view, alternations in sensory modality constitute a principled way to probe history-dependent cross-sensory dynamics in reaction time tasks. Recent works in the audio-visual domain have shown an increase in reaction times when the sensory modality of incoming targets switches (Crosse et al., 2022; Cuppini et al., 2020; Shaw et al., 2020; Vanneau et al., 2025), establishing the modality switch effect (MSE). In a preliminary study (Di Rosa et al., 2025), we repeated the behavioural experiment described by Crosse et al. (2022) using visuo-tactile stimuli,

List of abbreviations: MSE, Modality switch effect; ISI, Inter-stimulus interval; BIVIP, Bilateral visuohaptic processing; V, visual; T, tactile; VT, visuo-tactile; RT, reaction time; Rp, repeat; Sw, switch.

* Corresponding author at: Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi," University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy.
E-mail address: eleonore.dirosa2@unibo.it (E.F. Di Rosa).

https://doi.org/10.1016/j.neunet.2026.108960
Received 10 September 2025; Received in revised form 17 March 2026; Accepted 6 April 2026
Available online 7 April 2026

and similarly found evidence of cross-sensory inhibitory mechanisms. In line with previous reports on unilateral sensory switching (Spence et al., 2001), we observed delayed unisensory processing of unilateral haptic or visual targets preceded by a 1000–1500-ms inter-stimulus interval (ISI) from a heteromodal target. Novel predictions formulated by the preliminary network presented in Di Di Rosa et al. (2025) reproduced the slower responses to contralateral but not ipsilateral haptic or visual targets that occurred regardless of the modality of the preceding target (Berlucchi, 1995; Fendrich et al., 2004). Crucially, the network in Di Rosa et al. (2025) was not probed to determine how visuo-tactile MSE applies under bilateral contexts. If MSE operates in a hemisphere-specific manner, modality-specific repeat and switch trials involving both hemispheres would exhibit no significant differences in reaction times at shorter intervals. Alternatively, if MSE extends across hemispheres, a switch cost should be observed. Bilateral processing of visuohaptic information is further complicated by the temporal interplay between transcallosal communication and intra-hemispheric cross-sensory inhibition.

According to pivotal studies on side switching (Berlucchi, 1995; Marzi et al., 1991), the physiological delay in the inter-hemispheric transfer of neural information may contribute to a explain a faster detection of consecutive targets appearing in the same hemifield with respect to the body midline, i.e. ipsilaterally, compared to detection of targets in the contralateral hemifield. Exogenous shifts of attention to an unexplored location may also inhibit responses to sensory targets presented to a previously cued location, so that subjects experience a weaker 'inhibition of return' when target location shifts bilaterally (Berlucchi, 2006; Posner et al., 1985; Tamè & Longo, 2015).

Despite being extensive, the available evidence remains heterogeneous. To our knowledge, no previous work has systematically examined the interaction between modality switch effects and bilateral attentional orienting within an explicit architectural framework. Therefore, we developed an explainable multi-layer neural network that models bilateral visual and tactile temporal processing streams, incorporating inhibitory connections to replicate intersensory and inter-hemispheric interactions between vision and touch. We then used the model to predict whether modality switch effects should arise under bilateral stimulation and how these relate to unilateral effects. To provide a preliminary test of the predictions, we additionally present results from a simple reaction time task in which participants responded to visual, tactile, and visuotactile targets presented on one or both hands, alternating between crossed and uncrossed hand positions, under long and short ISI conditions.

To further probe the underlying mechanisms, we systematically perturbed the strength of the connections between layers in order to identify the processes responsible for bilateral visuo-haptic switch costs and the inter-individual variability observed in cross-sensory conflicts. Moreover, because vision can exert delayed influences on haptic detection accuracy (Cuppini et al., 2022; Lloyd et al., 2008; Mirams et al., 2017; Wani et al., 2021), we challenged the network with existing results from a four-alternative forced-choice task, asking whether intersensory switching extends beyond immediate stimulus-driven responses to shape haptic perceptual decisions over longer timescales.

Our approach provide a novel mechanistic insight into the neurobehavioral constraints of visuohaptic competition, proposing a plausible mechanistic description of how bilateral interactions shape multisensory processing and motor responses implicated in natural environments.

2. Material and methods

2.1. BIVIP network: qualitative description

The *Bilateral Visuohaptic Processing* (BIVIP) network stems from network models of visuo-tactile perception (Di Rosa et al., 2025, Cuppini et al., 2022; Magosso et al., 2010), incorporating additional visuo-tactile competition across hemispheres. It is composed of two symmetrical sub-networks, each mimicking visual and tactile processing in a single hemisphere. Each computational element represents a neuron pool with matching receptive fields in the brain, which may send and/or receive synaptic connections from other elements in the network. Synaptic weights represent a measure of overall connectivity effectiveness. Each simulated hemisphere includes two unisensory elements, one visual and one tactile (respectively V and T in Fig. 1), which are activated by external stimuli of the corresponding modality and aggregate unisensory information prior to integration in higher-order associative areas.

The visual and tactile units exchange cross-modal connections, Wtv and Wvt in Fig. 1, with fast dynamics responsible for visuo-haptic enhancement. Biologically, these connections mimic the cross-modal modulatory interactions between visual and somatosensory neural areas and promote integration of subthreshold visual and haptic stimuli if they occur in spatiotemporal congruency (Amedi et al., 2002; Cappe & Barone, 2005; Lucan et al., 2010; Merabet et al., 2008; Ramos-Estebanez et al., 2007).

The modality-specific visual and tactile interneurons, Iv and It respectively, are responsible for the mechanism of cross-sensory competition in that they are activated by direct excitatory connections, WIv and Wit, from modality-specific input units and they inhibit the activity of heteromodal ipsilateral and contralateral input units after competing through a winner-takes-all dynamic. Evidence of bilateral sensory suppression following heteromodal stimulation has been found in both animals and humans, with reports of direct long-range inhibitory cortico-cortical projections between low-level sensory cortices (Dinh et al., 2024; Weiler et al., 2024).

Long-range excitatory projections (Wmv and Wmt in Fig.1) connect the unisensory elements with a multisensory associative element responsive to vision and touch, mimicking the associative function of bimodal neurons in the superior parietal lobule (Brodmann areas 5 and 7, see Kandel, 2013). Indeed, their tactile receptive fields span the entire hand, while the matching visual receptive fields extend to the neighboring space around the skin.

Finally, each multisensory element can activate via interhemispheric excitatory synapses (WIntm in Fig.1) onto an interneuron in the contralateral hemisphere, which in turn inhibits the activity of ipsilateral unisensory elements. This top-down inhibitory architecture captures multiple bilateral visuotactile interactions (Sherrick, 1964; Tamè et al., 2011; Verrillo et al., 1983), and reflects a top-down feedback modulation from fronto-parietal higher-order circuits on sensorimotor regions involved in bilateral tasks (Macaluso et al., 2000; Macaluso & Driver, 2005; Rockland & Ojima, 2003).

2.2. BIVIP network: mathematical description

In the following section, each element will be indicated with a superscript, r (r = t, v, m, it, iv, or int, where 't' refers to the tactile input element, 'v' to the visual input element, 'm' to the multisensory element, 'it' and 'iv' to the tactile and visual interneurons, respectively, and 'int' to the top-level interneurons). Since this model is composed of two symmetrical networks, representing the left and right hemisphere, we will now provide a unique mathematical description of the network in one hemisphere, which is applicable to the other hemisphere. $u(t)$ and $y(t)$ represent the net input and output of a given neural unit at time t, respectively. Thus, $y^r(t)$ represents the output of the neural unit representing element r, described by the following differential equation:

$$\tau^r \frac{dy^r(t)}{dt} = -y^r(t) + F(u^r(t)) \quad (1)$$

$$F(u^r) = \frac{1}{1 + e^{-s(u^r - \theta)}} \quad (2)$$

where τ is the time constant, assigned in accordance with literature on neural membrane activation (Dayan & Abbott, 2001; Treves, 1993), and $F(u)$ represents the sigmoidal relationship. s and θ establish the slope

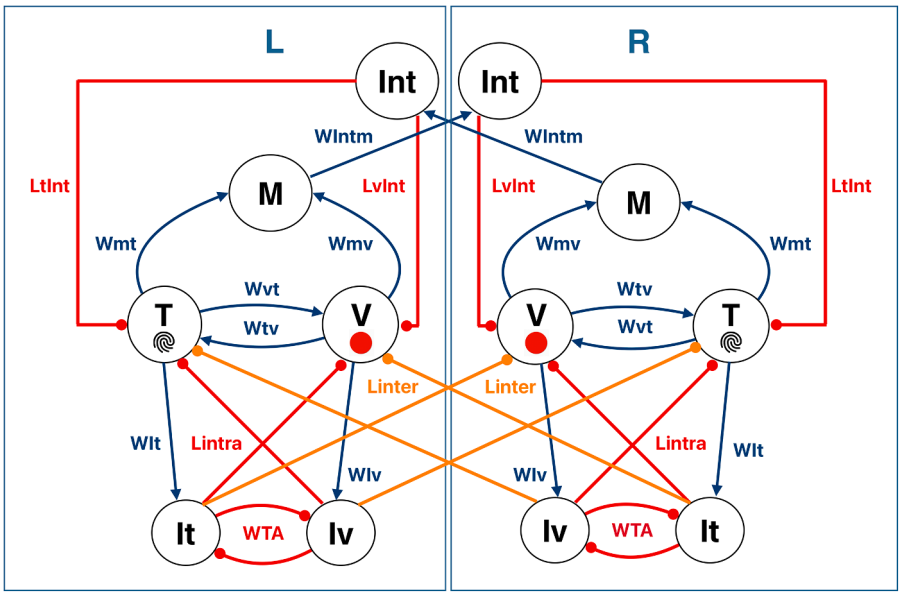


Fig. 1. Network architecture. Each of the two sub-networks (L for Left hemisphere, R for Right hemisphere) includes two input elements, one visual, V, and one tactile T. Elements V and T receive unisensory inputs, exchange cross-modal excitatory connections, Wvt and Wtv, and send feedforward excitatory projections (WIv and WIt) to modality-specific visual and tactile interneurons, Iv and It, which implement the cross-sensory inhibitory loop. Interneurons It and Iv compete via a winner-takes-all mechanism, so that the most activated modality may inhibit the heteromodal elements bilaterally, via feedback inhibitory synapses (Lintra - ipsilateral inhibition, Linter - contralateral inhibition). Excitatory projections Wmv and Wmt connect elements V and T with a multisensory element, M, in which incoming visual and tactile information is combined. Element M is connected to a top-down contralateral interneuron, Int, capable of inhibiting unisensory contralateral activity. Excitatory connections are in blue, inhibitory intra-hemispheric connections are in red, and inhibitory inter-hemispheric connections are in orange.

and the central position of the sigmoidal relationship, respectively. All neural units in the model share the same values of s and θ, chosen to ensure no residual activity for null input and a gradually varying neural response to increasing/decreasing external stimuli.

The net input to elements V and T is

$$i^r(t) = e^r(t) + c^r(t) - Li^r(t) - Lint^r(t),\ r = v,\ t \tag{3}$$

where (1) $e^r(t)$ represents the modality-specific external input, simulating the corresponding target, visual for unit V, and tactile for unit T; (2) $c^r(t)$ corresponds to the cross-modal component, delivered from the other-modality unisensory element through the cross-modal connections Wvt and Wtv; (3) the inhibitory interneural component, $Li^r(t)$, is generated by the other-modality interneuron in the same and in the opposite hemisphere; (4) the top-down inhibitory component, $Lint^r(t)$, is generated by the ipsilateral top interneuron which is excited by the contralateral multisensory element. The external input $e^r(t)$ has effectiveness I_0^r, set to avoid saturation of the sigmoidal function in the unisensory elements. and a duration T^r. Assuming a stimulus of sensory modality r (r = t, v):

$$e^r(t) = \begin{cases} I_0^r,\ 0 \le t \le T^r \\ 0,\ t \ge T^r \end{cases} \tag{4}$$

The time constants τ_v^e and τ_e^t have been set considering that the time needed for the activity in the sensory elements to reach 90 % of its steady state value is twice the time constant value. Neurophysiological studies show that the activation latencies of neurons in the striate cortex are typically longer (45–50 ms to elicit a detectable response after stimulus onset, see Maunsell & Gibson, 1992) than those of neurons in the primary somatosensory cortex (25–30 ms, see Chivukula et al., 2021; Reed et al., 2010). Therefore, we modeled external visual inputs with slower dynamics ($\tau_e^v = 25$ ms), than the one simulating the tactile input dynamics ($\tau_e^t = 15$ ms). Importantly, only the input parameters assume different values for the visual and tactile sensory pathways. All others are assumed symmetric (i.e. equal for both pathways) to reduce the number of assumptions. The cross-modal input, $c^r(t)$, is computed considering the activity of the other-modality element in the same hemisphere, amplified by the synaptic weight of the cross-modal synapses, which is identical for the two modalities ($W_{cross} = W_{tv} = W_{vt}$). Hence:

$$c^r(t) = W_{cross} \cdot y^s(t - \Delta t),\ r = v,\ t,\ s = v, t \tag{5}$$

where Δt is the latency in the exchange of cross-modal inputs between the two units. The time constant ($\tau_c^r = 15$ ms) and the delay ($\Delta t_c = 16$ ms) have been selected to reproduce electrophysiological findings (Ioannides et al., 2013; Schürmann et al., 2002), showing low-level tactile evoked visual activity as early as 50–60 ms after somatosensory cortical activation. The feedback inhibitory component, $Li^r(t)$, is derived from the other-modality ipsilateral and contralateral interneurons, and is defined as:

$$Li^{v(or\ t)}(t) = L_{intra} \cdot y^{it(or\ iv)\_ipsi}(t) + L_{inter} \cdot y^{it(or\ iv)\_contra}(t) \tag{6}$$

where $y^{it\_ipsi}(t)$ and $y^{iv\_ipsi}(t)$ are the activities of the presynaptic tactile and visual interneurons in the ipsilateral hemisphere as the considered unisensory element, and $y^{it\_contra}(t)$ and $y^{iv\_contra}(t)$ are the activities of the presynaptic tactile and visual interneurons in the contralateral hemisphere as the considered unisensory element. The effectiveness of intra-hemisphere feedback synapses, L_{intra}, is identical whether they originate in the visual interneurons and target the tactile element or originate in the tactile interneurons and target the visual element. The same holds for the inter-hemispherical feedback synapses, L_{inter}. As in Cuppini et al., 2020 and Di Rosa et al., 2025, we set the time constant of this mechanism to $\tau_l^r = 180$ ms to capture the empirically observed slow decay of cross-sensory suppression, which is strongest at ISIs around 1000 ms and persists over longer delays (Crosse et al., 2022). Following the initial activation of the unisensory input units and the subsequent recruitment of interneurons via excitatory projections (within the first ~100–300 ms), the feedback inhibitory component reaches its maximal influence after approximately three time constants (~540 ms), yielding

a substantial inhibitory impact on the other modality's unisensory units within an overall timescale of about 1 s. A further delay $\Delta t_{inter} = 8$ ms was also considered to model the inhibitory contributions originating in the contralateral hemisphere. The top-down inhibitory component, $Lint^r(t)$, is generated by the ipsilateral top-level interneuron and targets both ipsilateral unisensory elements. It is defined as:

$$Lint^r(t) = L_{Int} \cdot y^{int}(t),\ r = t,\ v \tag{7}$$

where $L_{Int} = LtInt = LvInt$ is the effectiveness of the top-down inhibitory synapses, and $y^{int}(t)$ is the activity of the top interneuron, in each hemisphere. In line with physiological studies (Hillis et al., 2006), the effectiveness has been set to enable the activation of both networks in response to bilateral stimulation, even for slight variations of stimulus intensity.

Each interneural unit (r = Ia, Iv) is excited by the unisensory unit of the corresponding modality via excitatory synapses with effectiveness $WI = W_{Iv} = W_{It}$, and exchanges inhibitory connections with the ipsilateral interneuron, implementing the WTA mechanism. Their net input is therefore:

$$i^r(t) = I^r_{ex}(t) - I^r_{in}(t),\ r = iv,\ it \tag{8}$$

It includes an excitatory component, $I^r_{ex}(t)$, and an inhibitory component, $I^r_{in}(t)$. For each interneuron, the excitatory components are defined as:

$$I^r_{ex}(t) = WI \cdot y^s(t),\ r = iv,\ it,\ s = v, t \tag{9}$$

where $y^v(t)$ and $y^t(t)$ are the activities of the presynaptic visual and tactile unisensory elements. These connections operate with a fast dynamic, hence $\tau^r_{I\_ex} = 15$ ms. On the other hand, the inhibitory component generated from the WTA mechanism is defined as:

$$I^r_{in}(t) = L_{WTA} \cdot y^s(t),\ r = iv,\ it,\ s = it,\ iv \tag{10}$$

where $y^{iv}(t)$ and $y^{it}(t)$ are the activities of the presynaptic visual and tactile interneurons, respectively, and L_{WTA} is the weight of the reciprocal inhibitory connections, identical for both interneurons in both hemispheres. The time constant has been set to $\tau^r_{I\_in} = 15$ ms.

The multisensory element receives a net input that is the sum of the stimuli, carried by long-range excitatory synapses, from the visual and tactile input elements. Hence,

$$ex^m(t) = \sum_r Wm_r \cdot y^r(t - \Delta t_m),\ r = v,\ t \tag{11}$$

where $Wm_r = Wmt = Wmv$ is the weight of the excitatory connections from the unisensory regions towards the multisensory element in each hemisphere. The time constant has been set to $\tau^r_m = 15$ ms to ensure a fast dynamic that allows to reproduce multisensory facilitation. The delay $\Delta t_m = 150$ ms agrees with the response threshold for fast outliers established during RT analyses.

The top interneuron in each hemisphere is excited exclusively by the contralateral multisensory element, so its net input is:

$$int(t) = W_{intm} \cdot y^{m\_contra}(t - \Delta t_{inter}) \tag{12}$$

where W_{intm} is the effectiveness of excitatory synapses to the top interneuron originating in the contralateral multisensory element, and has been chosen to facilitate competition between the two hemispheres for attentional resources within peripersonal space (Mattingley et al., 1997). $\Delta t_{inter} = 8$ ms is the temporal delay which accounts for the transmission of sensorimotor information across the corpus callosum (Fendrich et al., 2004). The time constant is $\tau^r_{Int} = 15$ ms, as other fast dynamics in the network.

As in previous works (Jansen & Rit, 1995, but see also Cuppini et al., 2014), synaptic dynamics are modeled with a second-order differential equation with two coincident real poles, implemented as

$$\begin{cases} \frac{d}{dt}o_i(t) = \delta_i(t); \\ \frac{d}{dt}\delta_i(t) = \frac{G^r_i}{(\tau^r_i)^2} i(t) - \frac{2\delta_i(t)}{\tau^r_i} - \frac{o_i(t)}{(\tau^r_i)^2}; \end{cases} \tag{13}$$

where $o_i(t)$ is the output of the differential equation for any generic input $i(t)$, as for example those described by any of Eqs. (4) to (7) for the input regions. As commonly done in computational implementations (Dayan & Abbott, 2001), $\delta_i(t)$ is used as an auxiliary variable, which allows to rewrite the second-order dynamics in first-order form, suitable for numerical integration. G^r_i is the gain and τ^r_i the time constant of the dynamics, for each region r. As reported in Table 1, G^r_i and τ^r_i have equal values for all connections, except for the external stimulation, and the cross-sensory feedback synapses.

2.3. Reaction time task: Experimental setup

2.3.1. Participants

6 healthy neurotypical individuals (4 males, mean age = 29.00 ± 6.89 years) were recruited to take part in the experiment. All participants gave their written informed consent before taking part and none of them reported any prior psychiatric, attentional, or developmental difficulties. The experiments were conducted in accordance with the Baylor College of Medicine protocols and the policies of the Institutional Review Board.

2.3.2. Stimuli and setup

All stimuli were generated using MATLAB Psychtoolbox-3 (r2022b, MathWorks). The experimental setup is displayed in Fig. 2A. The visual stimulus (V), adapted from Cuppini et al., 2020, consisted of a red disc (diameter = 3.2 cm) that flashed 15.5 cm to the left or the right and 0.4 cm above of a yellow central fixation cross, presented on a black background using a Dell monitor (resolution: 1280×1024 pixels). The tactile stimulus (T) was generated by passing an analog sinusoidal signal (100 Hz amplitude-modulated, carrier frequency = 200 Hz, sampling frequency = 48 kHz) through a 2 × 40 W stereo power amplifier (Pyle PTA2). The signal was delivered either to the left or right index finger via C-2 tactors (Engineering Acoustics Inc), fastened using self-adhesive CoFlex bandages and tape. The visuo-tactile stimulus (VT) consisted of the simultaneous presentation of the V and T stimuli on the same side, either to the left or right of the participant. All stimuli lasted 60 ms. Each participant was seated in a dimly lit room at a viewing distance of 70 cm from the monitor, with their arms resting comfortably on cushions in an outstretched position. To minimize external auditory interference, they wore isolating earmuffs over noise cancelling earphones (Tozo model

Table 1
Effectiveness of the network parameters.

Network parameters			
Neural units			
θ= 25	s = 0.3		τ = 3 ms
Inputs			
G^r_i= 75; r = t, v	τ^r_i = 15 ms; i = c, I_ex, I_in, m, Int	τ^r_e = 15 ms;	τ^y_e = 25 ms;
	T^t = 60 ms	I^v_0= [1.6 2.32]	T^v = 60 ms
Inhibition			
G^r_I= 750; r = t, v	τ^r_I = 180 ms; r = t, v	τ^{Int}_r = 20 ms; r = t, v	
Synaptic connections			
Wcross = 0.3			Δt_c = 16 ms
WI= 2	$L_{intra} = L_{inter}$ = 0.06		L_{WTA} = 3
Wm = 2.2			Δt_m = 150 ms
WIntm = 0.6	LInt = LtInt = LvInt = 0.2		Δt_{inter} = 8 ms

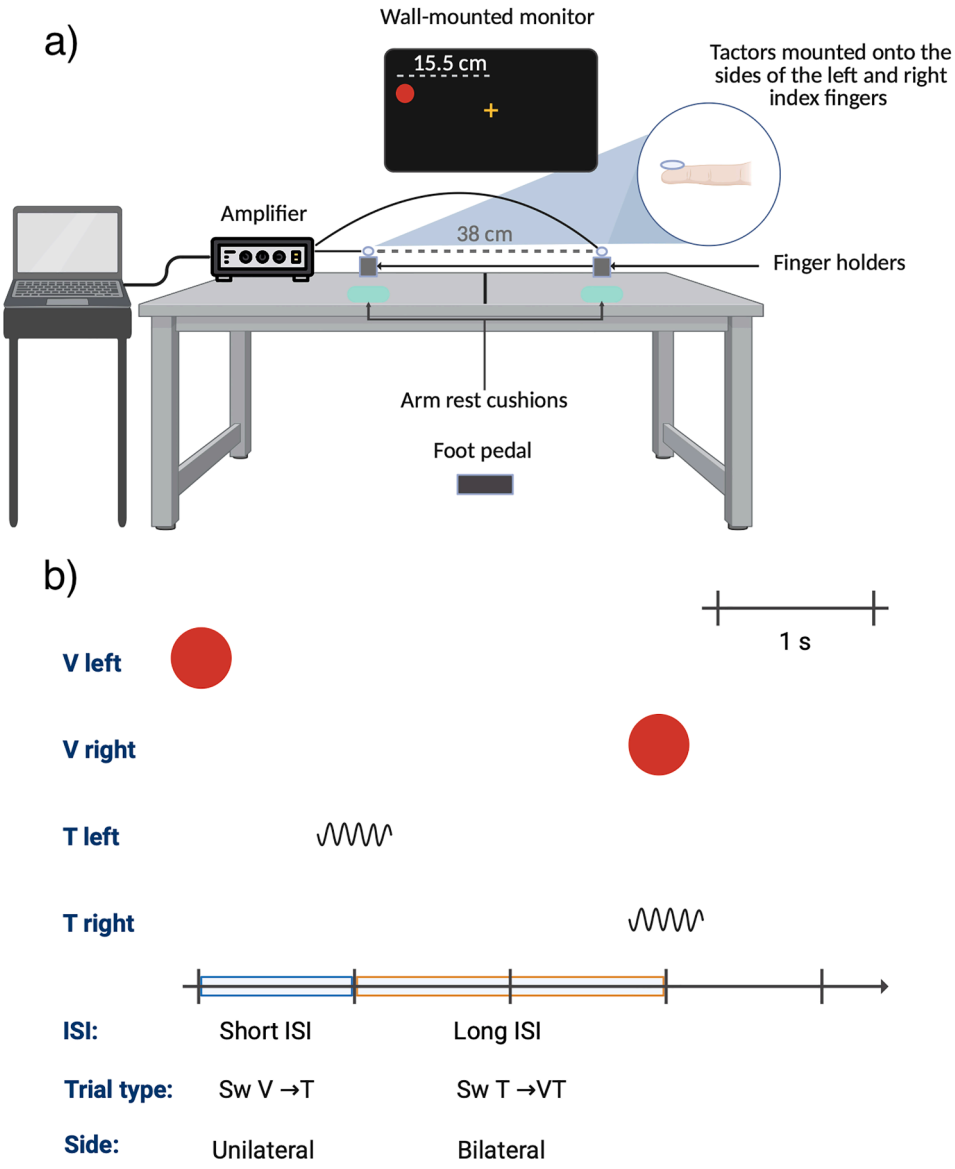


Fig. 2. Experimental setup and example trial representation. a) Participants maintained their gaze on a yellow fixation cross while attending to stimuli and press on a foot pedal as soon as they detected a stimulus. The visual stimuli were presented on the monitor, and the tactile signals were delivered through tactors after amplification. The participants maintained their hands on two cushions and placed their index fingers, strapped to tactors, on finger holders spaced at 19 cm from the midline on each side. b) The experiment was conducted in two arm positions that alternated between blocks: crossed and uncrossed. Visual, tactile, or visuo tactile stimuli were randomly presented to the left or right of the midline of the participant.

NC9) that played white noise. Their index fingers were placed on finger holders at a distance of 38 cm from each other. Reaction times (RTs) were measured as foot pedal presses upon detecting a stimulus. The stimuli were presented in 8 blocks of 150 trials each with a 5-minute break between blocks. RTs were measured under two experimental conditions: arms crossed and uncrossed across the midline of the participant, alternated between blocks, with the starting condition counterbalanced across participants. The experimental blocks were preceded by two training blocks of 24 trials conducted in both arm positions. The stimuli were presented in a random order with equal probability of occurring to minimize response bias. As in Crosse et al. (2022), ISI varied randomly between 1000–3000 ms according to a uniform distribution.

2.4. Reaction time task: Data analysis

RTs were analyzed using MATLAB. Figures were obtained in Python. RTs were excluded from further data analysis if they were shorter than 100 ms (anticipatory responses), longer than 2 s (slow responses), or misses. In case of multiple pedal presses the first response made after 100 ms after the onset of a stimulus was collected. Furthermore, the code did not collect 'false presses' or responses made before the onset of a stimulus in a given trial. Trials were categorized by four factors: (1) ISI, whereby those with ISIs $\langle$ 1.5 s were labeled as 'short' whereas those with ISIs $>$ 2.5 s as 'long'; (2) presentation side, defined as 'unilateral' when consecutive targets were presented on the same side and 'bilateral' when consecutive targets were presented on different sides; (3) modality type, as V, T or VT based on the modality of the following stimulus; and (4) trial type, such that modality repeats (Rp) were consecutive trials in which stimuli shared the same sensory modality, whereas modality switches (Sw) involved a change in modality. Switch costs were computed as the difference between the mean RTs for Sw trials and Rp trials.

Data distributions were tested for normality using the Shapiro-Wilk

test, and when found to violate this assumption non-parametric tests were implemented. Friedman tests with quantification of Kendall's W were applied to assess the effects of arm position (crossed, uncrossed), modality (V, T, VT), and trial type (Rp, Sw) on RTs, to explore the emergence of modality switch effects in the visuotactile condition. To determine whether modality switch effects were modulated by ISI and presentation side, 2-way repeated-measures ANOVA were implemented with factors trial type (Rp, Sw), and presentation side (unilateral, bilateral) respectively on unisensory short ISI RTs, and on unisensory long ISI RTs, and with factors ISI (short, long) and trial type (Rp, Sw) respectively on unilateral RTs, and on bilateral RTs. Significant effects were followed by post hoc Tukey HSD - corrected pairwise comparisons.

2.5. Assessment of network performance

First, we verified that the network could reliably simulate unilateral intersensory switching between vision and touch by replicating experimental switch costs resulting from the 'divided attention' block in Spence et al., 2001. Specifically, we chose this study because it contained visual and tactile stimulation, all stimuli were treated as targets, and were separated by ISIs of 1.5 – 1.9 s, which corresponds to the average standard range chosen in studies of audiovisual switching (e.g. Crosse et al., 2022). Using the same ISI range, we collected mean RTs over 150 network simulations for Sw and Rp conditions (unisensory V and T), with sensory inputs chosen from a uniform distribution (see Table 1), and computed the switch costs with respect to each modality.

To test the network's ability to predict the temporal outcome of bilateral interactions, we simulated randomized sequences of unilateral and bilateral V, T, and VT inputs, with pseudorandom ISIs ranging between 1000 and 3000 ms according to a uniform distribution. Simulated RTs were obtained as the time difference between the onset of each stimulus and the first time point during which the activation of the multisensory element exceeded 30 % of the maximum level. Each input intensity was chosen from a uniform distribution to mimic within-subject variability among trials. Simulated RTs were classified as empirical RTs. For each simulated participant, mean RTs per condition were computed over 150 simulations with random input value, and simulated switch cost as the difference between RTs to Sw and to Rp trials, for each modality type, side, and ISI condition. Empirical and predicted RTs were quantitatively compared using nonparametric Mann–Whitney U tests, separately for each combination of trial type (repeat vs switch), modality type (V, T, VT), and presentation side (unilateral vs bilateral), and by computing Root Mean Square Error (RMSE) for each condition. To better characterize the behavior of the network in bilateral conditions at short timescales, an additional set of simulations was also implemented, for ISI ranging between 50 and 300 ms, and the corresponding results are shown in the Supplementary Material.

We then performed a sensitivity analysis to elucidate the plausible neural mechanisms underlying visuo-tactile switching, by randomly perturbing the effectiveness of the synaptic connections specifically involved in bilateral interactions (i.e. Wm, Linter, LInt). For each of these connections, we progressively increased the noise applied to the basal effectiveness in three steps (i.e. 10 %, 50 % and 100 % of noise with respect to the basal effectiveness), generating three uniform distributions of synaptic effectiveness of increasing variance. From each distribution, 6 different values were chosen to simulate 6 different participants, and used to obtain simulated RTs in 8 blocks of 50 repetitions per unisensory trial type. For each synaptic distribution, we analyzed the dispersion of the joint RT distributions of Rp and Sw RTs (simulated and empirical), by decomposing them into a diagonal component $s = (Rp + Sw)/\sqrt{2}$, which allows to capture variability common to both trial types, that is overall RT fluctuations, and an orthogonal component $d = (Sw - Rp)/\sqrt{2}$, highlighting variability specifically related to the deviation between repeat and switch responses, i.e. switch costs. We then quantified the parallel ($M_{\parallel}$) and orthogonal ($M_{\perp}$) mismatch between the empirical (subscript 'sub') and the simulated (subscript 'net') distribution in each synaptic configuration, using the log-ratio of variances along the axes defined by components s and d, respectively.

$$M_{\parallel} = \left| log \frac{var(s_{net})}{var(s_{sub})} \right| M_{\perp} = \left| log \frac{var(d_{net})}{var(d_{sub})} \right|$$

The log-ratio provides a scale-independent measure of mismatch, so that equal proportional increases and decreases in variability produce the same magnitude, and a value of zero corresponds to identical dispersion. Greater values of mismatch are indicative of a weaker fit between empirical and simulated results.

Finally, we simulated the 4-alternative forced choice (4AFC) task presented in Wani et al., 2021 to explore possible effects of MSE on visuo-tactile detection accuracy. The task comprised two blocks: a T block, in which participants were required to detect a tactile target delivered on one of their two hands (unimanual), both hands simultaneously (bimanual) or none; and a VT block, in which a bright LED, located on one hand, or a dim LED, on the other hand were lit alone or together, generating visual cues paired with tactile stimulation on one, two or no hands. In all, the T block included four conditions, i.e. tactile target on one hand (left/right), two hands or none, while the VT block combined the four tactile conditions with four visual conditions, i.e. bright LED, dim LED, both LEDs, no LED, i.e. 16 VT conditions overall. In the T block, unimanual targets were detected more accurately (78 ± 2 %) than bimanual targets (68 ± 4 %), while results from the VT block revealed a systematic effect of concurrent visual stimulation on tactile performance, with decreased detection accuracy even in the no-LED condition. During the VT block, the tactile-only no-LED stimulation could be presented following a trial comprising bright-, dim-, both-LED stimulation with an inter trial interval of approximately 1 s, where the unisensory switch cost is expected to reach its peak. For this reason, we simulated the experiment with our network, and classified trial response accuracy based on the modality of the preceding trial, to seek whether modality switching could play a role in the worsening of tactile detection. The effectiveness of the tactile and visual inputs was adapted in accordance with the sensory inputs used in Wani et al. (2021), and a noisy component was added to the inputs to mimic inter-trial variations.

3. Results

3.1. Unilateral network performance

Fig. 3a compares mean switch costs, computed as the difference between Sw and Rp trials, to unilateral visual and tactile targets reported in Spence et al., 2001, indicated as 'Empirical - V' and 'Empirical – T' respectively, and their simulated counterpart. Though no significant differences were found between empirical and simulated distributions, the network seemed to slightly underestimate mean cost values for both modalities. Reasons for this observation may include the presence of auditory targets, alongside visual and tactile, within the experimental paradigm used by Spence et al., 2001, which may reduce the predictability of the incoming stimulation for the participants thus leading to an increase in sensory perceptual workload and thus in behavioral costs overall.

3.2. Bilateral network validation: Experimental data vs network predictions

Fig. 3 also shows experimental and predicted switch costs for short and for long ISIs, separately for unilateral and bilateral presentation, separated by modality type. Both unilateral and bilateral unisensory stimulations yielded greater switch costs than the cross-modal condition, though a significant difference emerged at short ISIs only for the

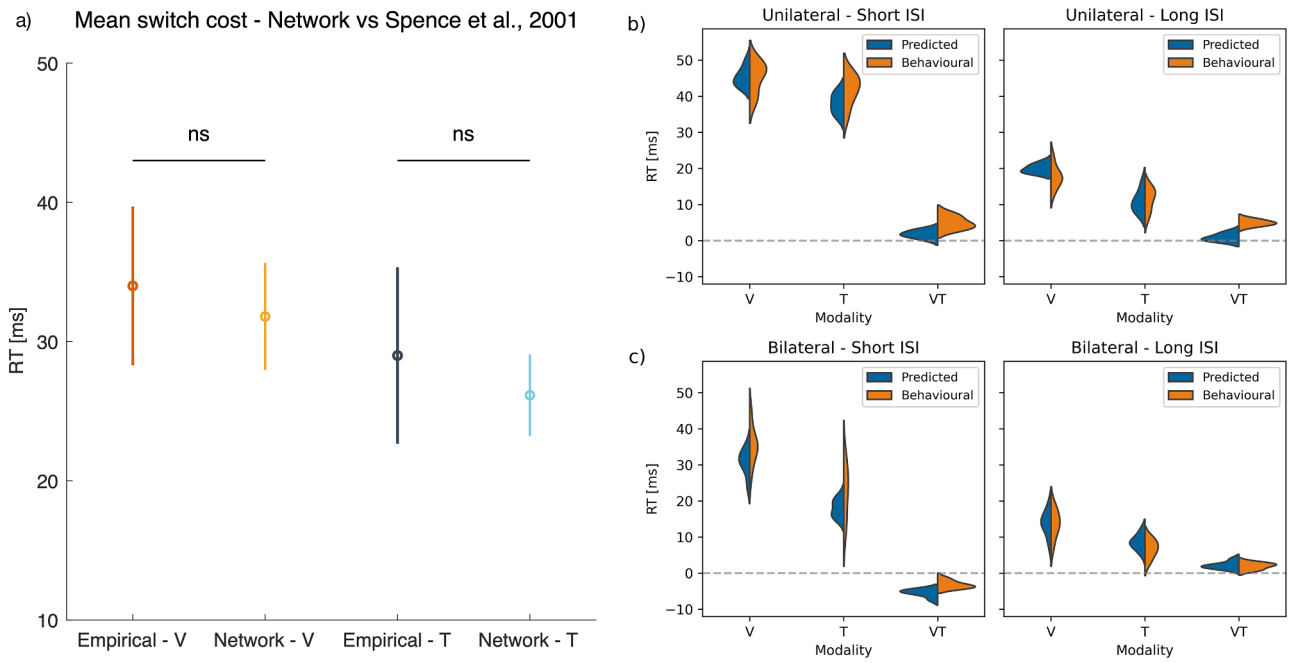


Fig. 3. Simulated mean switch cost for unilateral presentation and predicted switch costs for unilateral and bilateral short and long ISI. Data vs Model. a) Mean-ISI unilateral switch cost (mean ± SEM) from Spence et al., 2001 and network simulations, for the visual and tactile modality. A 1-way ANOVA confirmed no main effect of source (empirical vs network) for both modalities (ns = nonsignificant). b) Switch Costs (mean ± SEM) per modality (V, T and VT) and per ISI (Short, left panel and Long, right panel) for unilateral target presentation. c) Switch Cost (mean ± SEM) c per modality (V, T and VT) and per ISI (Short, left panel and Long, right panel) for bilateral target presentation.

unilateral condition (unilateral $\Delta = 0.028$, 95 % CI = [−0.004 0.061], $p < 0.05$, *bilateral:* $\Delta = 0.012$, 95 % CI = [1.0e-4 0.025], $p = 0.051$). In line with Di Rosa et al. (2025), these unilateral results support the idea that modality switch effects occur between vision and touch, which are maximal at short ISI and reduced at longer ISIs. Though lower than in the unilateral conditions, both empirical and predicted switch costs are observed when targets were delivered bilaterally. These are greater at short ISIs than at long ISIs, thus suggesting that MSE may extend to bilateral visuo-haptic interactions. Notably, the VT switch cost in bilateral conditions was negative, thus signaling longer RTs to Rp as compared to Sw trials. No significant main effect of source (model vs experimental) was found (all p > 0.1). ISI ($p = 0.0253$, *Kendall* $W = 1.00$) and modality ($p = 0.0067$, *Kendall* $W = 1.00$) had main effects on RTs, with post hoc comparisons showing that RTs were significantly faster in the VT condition compared to both the T condition (p = 0.046) and the V condition ($p = 0.008$). No significant main effect of arm position was observed ($p = 0.180$, *Kendall* $W = 0.360$), consistent with previous reports indicating that crossing arms did not induce any significant changes in the dynamics of visuohaptic intersensory switching (Heed & Azañón, 2014). RMSE did not exceed 10 ms in any of the 24 conditions. The greatest discrepancy was found in the SwT (i.e. switch V to T) – Short ISI – Unilateral condition (meanRMSE ± SEM = 8.51 ± 1.42 ms). On average, the model performed slightly better in the V (3.63 ± 0.42 ms) than in the T (5.00 ± 1.03 ms) and VT conditions (4.45 ± 0.98 ms), and in Rp (3.84 ± 0.73 ms) than in Sw trials (4.87 ± 0.84 ms). Results were comparable for Short (4.63 ± 1.01 ms) vs Long ISI (4.04 ± 1.0 ms), and for Unilateral (4.69 ± 0.94 ms) vs Bilateral presentation (4.03 ± 0.98 ms). Fig. 4 shows experimentally obtained RTs at short and long ISIs. For each ISI condition, we report RT values to repeat and switch trials, when the targets appear on the same hand or on different hands. At short ISIs (Panel a), RTs to Sw trials were slower than those to Rp trials ($F_{1/5} = 20.10$, $p = 0.01$), both in case of unilateral presentation (*median ± SEM, sw: 370.20 ± 2.00 ms, rp: 323.60 ± 2.31 ms,* $\Delta = 0.039 \pm 0.016$, $p < 0.01$, *95 % CI = [−0.005 0.084],* bilateral: *rp: 340.62 ± 4.16 ms, sw: 358.27 ± 4.44 ms,* $\Delta = 0.019 \pm 0.04$, $p < 0.01$, *95 % CI = [0.009 0.030)*. Presentation side showed a main effect ($F_{1/5} = 46.59$, $p = 0.002$), and unilateral repeat led to faster, though not significant RTs than bilateral repeat ($p = 0.46$). On the contrary, switch trials led to

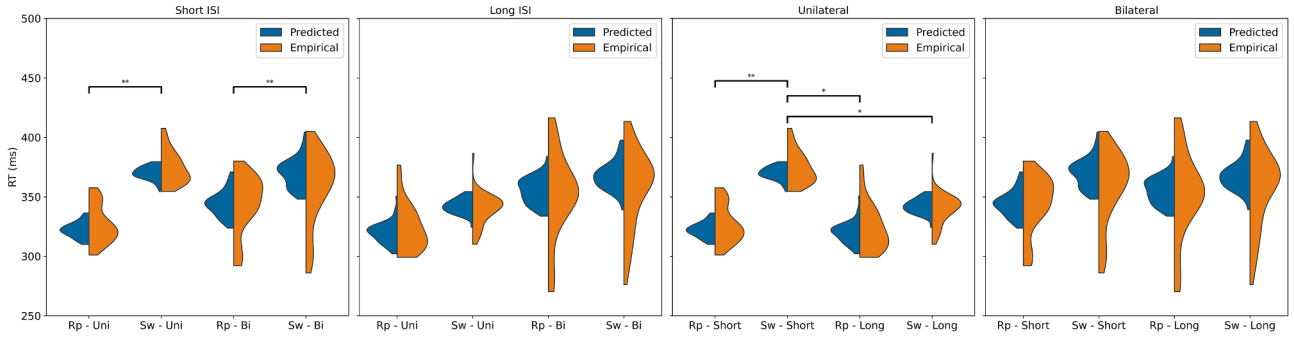


Fig. 4. Effect of trial type and cue location on mean RTs. a) RTs at short ISI in Repeat and Switch conditions for stimuli presented on one hand and on both hands. b) RTs at long ISI in Repeat and Switch conditions for stimuli presented on one hand only and on both hands. c) RTs to unilateral presentation of Rp/Sw trials at short/long ISI d) RTs to bilateral presentation of Rp/Sw trials at short/long ISI.

slower RTs when the target was delivered unilaterally, but again significance was not reached (p = 0.36). At long ISIs (Fig. 4b), median RTs were generally slower for Sw than for Rp trials. Despite not reaching significance, median switch RTs for bilateral stimulation were slightly greater compared to unilateral stimulation RTs (*sw bi: 352.96 ± 4.52 ms* vs *sw uni: 342.5 ± 1.86 ms*). Finally, Fig. 4c outlines an additional effect of ISI for unilateral presentation ($F_{1/5} = 21.68$, p = 0.009, in switch trials $\Delta = 0.06 \pm 0.02$, $p < 0.05$), and that Sw trials at short ISI led to slower RTs than Rp trials at long ISI in case of unilateral presentation ($\Delta = 0.034 \pm 0.008$, $p < 0.05$, *95 % CI = [0.012 0.056]*).

3.3. Sensitivity analyses

3.3.1. Bilateral unisensory conditions

To test whether perturbations of the network mechanisms could account for within and/or between-subjects variations in the empirical results, we introduced artificial variability in the effectiveness of the synaptic connections included in the model. First, we focused on RTs to unisensory bilateral trials, which resulted mainly dependent on the effectiveness of (1) the excitatory feedforward connections towards the multisensory element, Wm (Fig. 5, panels a), b) and c)), and (2) the inhibitory feedback connections from the sensory interneurons to the contralateral unisensory elements, Linter (Fig. 5, panels d), e) and f)). For increasing variability of the sole feedforward connections, simulated

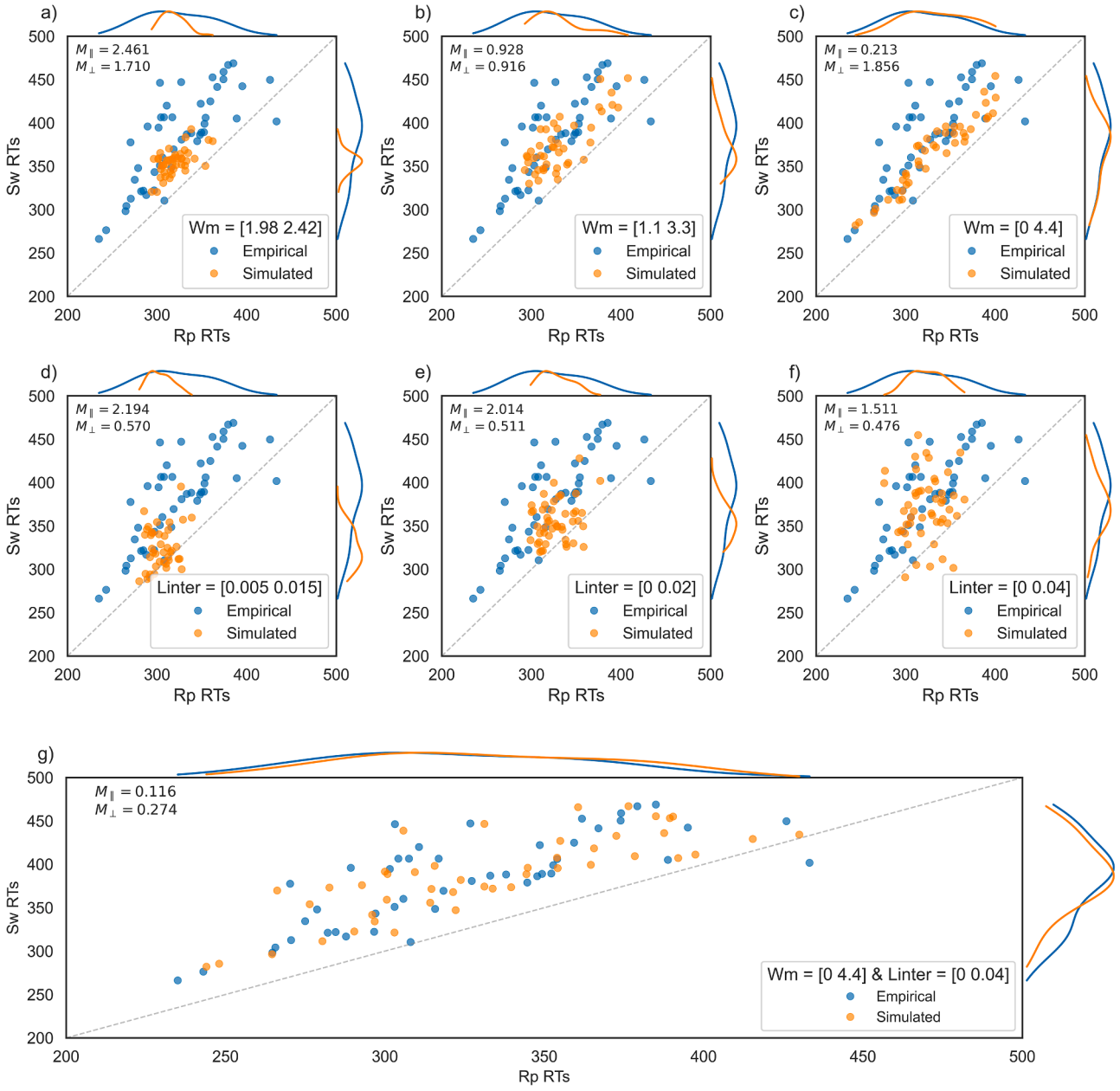


Fig. 5. Sensitivity analyses on unisensory bilateral interactions. Empirical and simulated Sw RTs vs Rp RTs displayed for increasing perturbation of the effectiveness of the feedforward connections, Wm (panels a to c), and of the feedback inhibitory connections to the input elements, Linter (panels d to f). Model results from panel g were obtained combining configurations c and f. Probability distribution functions of Rp RTs and Sw RTs are reported above and to the right, respectively. Each blue dot represent the mean empirical RT over a block of ~50 trials, for each participant. Orange dots represent corresponding results from model simulations. The values of the synaptic connections in each configuration were chosen from a uniform distribution, whose limits are presented in square brackets in the legend.

RTs show a greater tendency to replicate variability in the distribution of empirical Rp data, as confirmed by the pdf profile above each graph and by a monotonic decrease in the parallel mismatch between the distributions. When only Linter was perturbed, the distribution of simulated RTs showed greater dispersion over the y axis, thus better capturing variations in the difference between Rp and Sw trials. Consequently, the pdf distribution of simulated Sw trials gradually aligns to the empirical pdf of Sw RTs, and the orthogonal mismatch between the distribution reduces consistently. Finally, we applied the maximal noise perturbation on both Wm and Linter, and we obtained a simulated distribution (in orange in Fig. 5, panel g), whose mismatch is comparatively lower than in the previous cases both along the parallel and orthogonal components. While only the joint contribution of the feedforward and the contralateral inhibitory mechanisms allowed the model to fully recapitulate the global variability in the empirical distribution, this analysis clearly highlighted separate roles for the two pathways: the former resulted mainly responsible for differences in temporal detection independent on trial type, while the latter was more supportive of the emergence of MSEs between vision and touch across hemispheres.

3.3.2. *Bilateral bisensory conditions*

While maintaining the maximal percentage of noise on Wm and Linter, we applied the same perturbation procedure to each of the remaining connections in the network and simulated Rp and Sw trials during VT stimulation. The strongest effect emerged by manipulating connections involved in the inter-hemispheric top-down inhibition, i.e. the excitatory connections from the multisensory elements towards the top interneurons, WIntm, and the inhibitory top-down feedback connections to the ipsilateral unisensory elements, LInt. More specifically, the inter-hemispheric top-down control depends on the product of these two synaptic weights. Fig. 6 shows, along panels a) to c), that varying Lint allowed to improve the similarity between empirical and simulated bisensory RT distributions. Both Rp and Sw pdfs show increasing adherence to the empirical curve, and both parallel and orthogonal mismatch decrease accordingly. Unlike data from unisensory trials, whose Rp-Sw RT pairs are located almost completely above the diagonal and are indicative of positive switch costs, in bisensory trials a slightly greater number of responses to Rp trials were slower than those for Sw trials, thus leading to a negative switch cost. These observations are in line with the switch cost data for VT bilateral stimulation (Fig. 3, panel c), especially at short ISIs.

3.4. *Tactile response accuracy in bilateral visuotactile tasks*

The network successfully recapitulates response accuracy to tactile stimulation (Fig. 7), especially in terms of hit rates and correct rejections, both in the case of unimanual (Bright associated/Dim-associated hand) and with bimanual stimulation (Both hands). Predicted accuracy values confirm a decline in tactile performance when tactile targets are interleaved with visual or visuotactile targets within the experimental block with all conditions exhibiting a reduction in hit rates as compared to the tactile-only block (Fig. 7, right column). Empirical data from the visuotactile block indicate higher accuracy for the Bright LED-associated hand compared to the Dim LED-associated hand (70 ± 5 % vs 64 ± 3 %), though this difference did not reach significance. In contrast, simulated hit rates are better for the Dim LED-associated hand (73 ± 4 % vs 76 ± 3 %), which seems to be thus less affected by intersensory switching effects.

4. Discussion

Interacting with a multisensory environment often requires processing visual and haptic signals in rapid succession, thus leading researchers to choose task-switching paradigms for veridical perceptual investigations. This intermixed stimulus presentation design gives rise to modality switch effects, which systematically contribute to integration by being typically larger on unisensory than on multisensory trials. For this reason, understanding visuo-haptic competition over time and space is crucial to decode the principles of perceptual integration and decision-making by correctly interpreting experimental results aimed at mimicking ecologically valid situations. Visuotactile interactions have been the subject of extensive research from a behavioral, neurophysiological, and computational perspective.

Specifically, computational modeling has been exploited to investigate the enhancement deriving from concurrent visuohaptic stimulation, with particular attention to the perceptual changes generated to variations in the spatial configuration between haptic and visual targets. For instance, Diederich & Colonius, 2007 exploited an extended version of the Time-Window-of-INtegration model (Colonius & Diederich, 2004), to investigate the facilitatory effect of concurrent vibrotactile stimulation on visual target detection and its modulation due to physical distance, eccentricity or frequency of tactile cues. While focusing more on the computation underlying these effects than on the potential neural substrates, the authors propose a rigorous and comprehensive characterization of visuotactile interactions across varying spatial configurations, providing evidence for bisensory enhancement extending to bilateral stimulation. More recently, visuo-tactile interpretable models have been developed to augment the sensing capabilities of robotic agents.Donato et al. (2024) introduced a generative visuotactile model for soft robots that leverages proprioceptive and visual inputs to predict

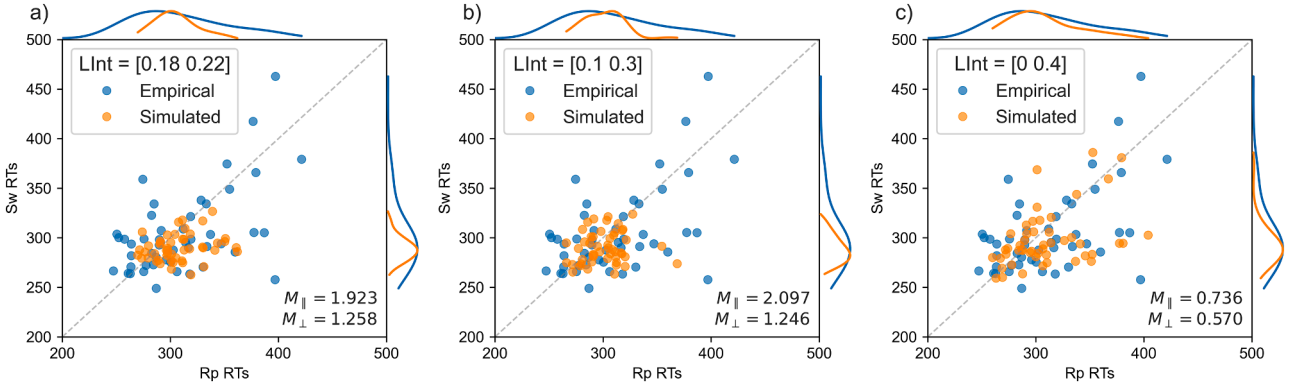


Fig. 6. Sensitivity analyses on bisensory bilateral interactions. Empirical and simulated Sw RTs vs Rp RTs displayed for increasing perturbations of top-down inhibitory connections to the input elements, Lint. Simulated RTs were obtained maintaining the values of feedforward and feedback inhibitory connections shown in Fig. 5, panel g). Probability distribution functions of Rp RTs and of Sw RTs are reported above and to the right, respectively. Each blue dot represent the mean empirical RT over a block of ~50 trials, for each participant. Orange dots represent corresponding results from model simulations. The values of the synaptic connections in each configuration were chosen from a uniform distribution, whose limits are presented in square brackets in the legend.

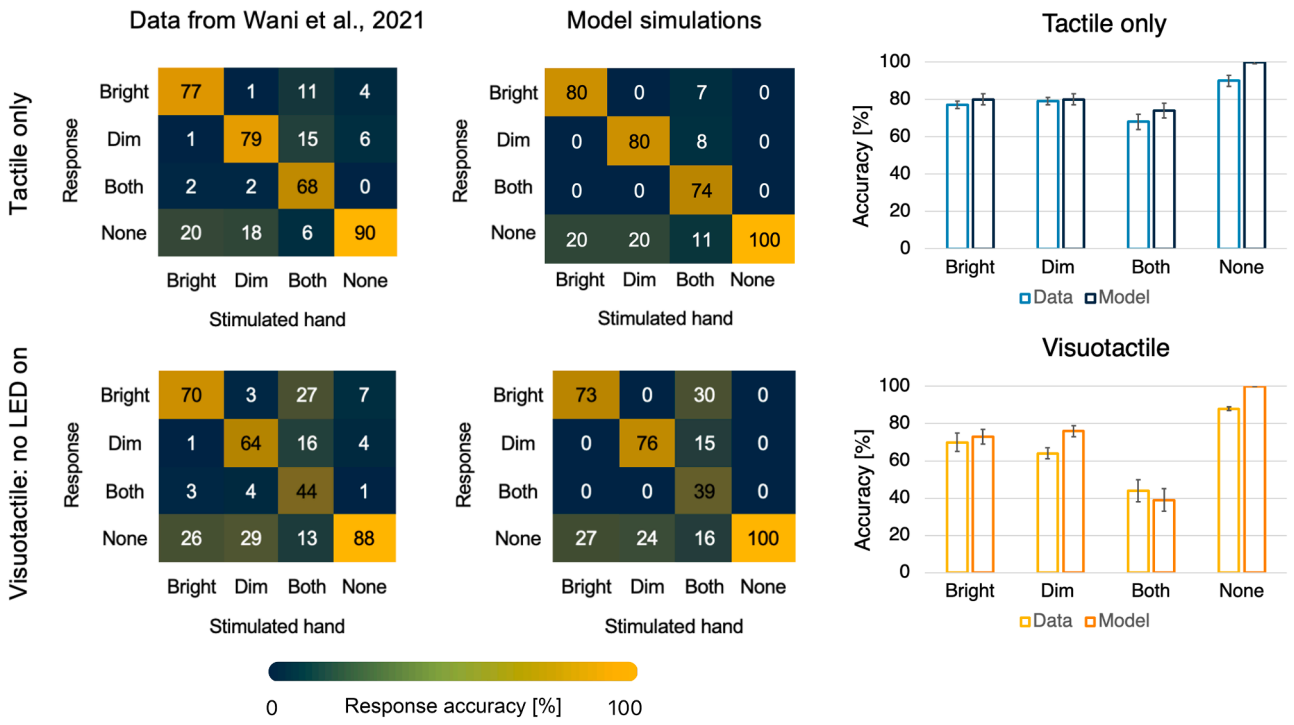


Fig. 7. Tactile response accuracy in bilateral interactions. Confusion matrices showing mean response accuracy on trials with tactile stimulation on the hand associated with the bright LED, the hand associated with the dim LED, both hands simultaneously or no tactile stimulation. Bar plots on the right represent data vs model simulated hit rates for the two conditions.

contact events. Their Conditional Variational Autoencoder, with modality-specific encoders, enables interpretable latent representations and cross-modal inference, enhancing spatial perception in compliant robotic systems. While their model primarily accounts for spatial encoding, our network unfolds the temporal structure of visuotactile processing in facilitatory and competitive contexts, offering novel theoretical contribution to achieve synchronized, human-aware sensorimotor integration in robotics.

The BIVIP network presented in this work builds on a pivotal study by Magosso et al. (2010), in which visuotactile spatial attention was investigated using a bilateral neural network reproducing both bisensory cooperative enhancement and inhibitory interactions between hemispheres in healthy and brain-damaged subjects. A variant of the same architecture was later exploited by Cuppini et al. (2022) to shed light on the neural correlates of the visually induced bias in bimanual spatial touch.

In the present network, we assumed that postural and proprioceptive information is implicitly incorporated within the tactile representation, rather than implemented as a separate input pathway. While we are aware that proprioceptive signals and posture are generally crucial for constructing body-centred spatial representations that support the localisation of external events, particularly under posture changes such as hand crossing, several considerations support our assumption within the temporal regime targeted here. Touch is initially encoded in a body-based reference frame, as reflected in the organisation of somatosensory cortices, whereas visual signals are initially represented in external coordinates. A substantial body of literature suggests that tactile localisation requires a remapping process that progressively aligns tactile representations with external reference frames, and that posture-related effects (e.g., hand crossing) are especially pronounced when tasks require resolving conflicts between anatomical and external spatial codes at short time scales. In particular, robust crossing effects have been reported for explicit spatial discrimination tasks and for tactile temporal judgments at very short stimulus onset asynchronies, often in the 50–300 ms range (Azañón et al., 2015; Badde et al., 2016; De Haan et al., 2012, but see also Heed & Azañón, 2014). Critically, these effects have also been shown to diminish when sufficient time is available for tactile remapping into external coordinates, and when task demands do not require explicit spatial discrimination, as in simple reaction time paradigms (Anzola et al., 1977; Craig & Belser, 2006; Nicoletti et al., 1984; Yamamoto & Kitazawa, 2001). In this context, the primary aim of the present study is to model dynamics that emerge at inter-stimulus intervals on the order of ~1 s and beyond, i.e., the temporal scale typically associated with modality switch effects (Crosse et al., 2022; Shaw et al., 2020; Spence et al., 2001). Because these ISIs are several times longer than the temporal window in which remapping-related posture effects are most strongly expressed, it is reasonable to assume that proprioceptive contributions may have comparatively less leverage on the specific switching dynamics investigated here, relative to the interaction between visual and haptic streams. For completeness, we performed an additional set of predictive simulations in bilateral conditions for all modalities and trial type combinations, using ISI between 50 and 300 ms, and reported the results in the supplementary material.

Second, in the behavioural context reproduced by the model, haptic detection occurs while participants visually explore space. Under these conditions, vision often provides highly reliable spatial information and can dominate spatial estimates when sensory cues are integrated, potentially reducing the behavioural footprint of posture-related signals in non-spatial detection tasks (Lateiner & Sainburg, 2003). Third, if posture-related signals were a dominant driver of performance in this paradigm, crossing the hands would be expected to introduce systematic detection delays. Conversely, if tactile remapping is largely settled at the ISIs relevant for modality switch effects and the task does not require explicit spatial discrimination, differences between crossed and uncrossed posture should be reduced, and any remaining effects would likely be small relative to the modality-driven switching dynamics.

Despite the extensive documentation of cross-sensory temporal competition in audiovisual processing, to our knowledge no previous work has explicitly addressed the temporal dynamics of sensory switching in the visuotactile domain, particularly in tasks involving

bilateral stimulus presentation. Furthermore, multi-layer neural network models providing comparable insight into the underlying neural processes specific to these dynamics are still lacking. By virtue of the interneural competitive loop between vision and touch, which constitutes the innovative core of the present architecture, BIVIP provides a plausible framework to reproduce perceptual outcomes of bilateral visuotactile interactions across a broad temporal axis, from near-synchronous stimulation to temporally offset cues spanning several seconds. Overall, this network allows to reproduce three neural mechanisms. First, multisensory facilitation is enacted at early processing stages: visual and tactile input layers reciprocally excite one another via cross-modal connections and jointly activate the multisensory layer through fast excitatory synapses. This mechanism enables the network to replicate the enhancement of perceptual responses to congruent visuotactile stimuli. Such integration is well documented and is thought to occur both through direct projections between early sensory cortices (e.g., S1–V1, S2–V5/MT), and via higher-level multisensory regions such as the posterior parietal cortex (PPC), intraparietal sulcus (IPS), and superior temporal sulcus (STS), which exhibit bimodal or trimodal responses and mediate cross-modal congruency effects (Gentile et al., 2011; Macaluso & Driver, 2005).

Second, cross-sensory competition, implemented by a feedback connectivity loop with slower dynamics, governs reciprocal inhibition between visual and tactile modalities, both within and across hemispheres. This circuit includes excitatory projections to interneurons (WIv and WIt) and inhibitory feedback synapses targeting the unisensory layers (LInter and LIntra). Our results support the notion that this mechanism is responsible for cross-modal suppression effects, whether stimuli are delivered to the same hemifield or across hemispheres, and may explain modality switch effects between vision and touch occurring unilaterally and bilaterally. Experimental data and network predictions showed significantly longer reaction times for bilateral unisensory switch trials compared to repeat trials, particularly at shorter inter-stimulus intervals, consistent with the canonical profile of intersensory switching reported in the audiovisual literature (Crosse et al., 2022; Shaw et al., 2020; Vanneau et al., 2025). Despite the overall good fit, the model tended to predict slightly smaller switch costs than those observed empirically, albeit without reaching statistical significance. One possible reason could be that the present architecture does not include an explicit auditory processing stream. Evidence from multisensory research shows that the presence of an additional modality can alter perceptual weighting and introduce further competition or facilitation, potentially modulating switching effects (Bresciani et al., 2008; Hagmann & Russo, 2016; Hecht & Reiner, 2009). As explained in the SubSection 2.3 of Material and Methods, the empirical results used for comparison with network predictions were collected under conditions that minimized auditory input, yet residual differences may still reflect the absence of tri-sensory interactions in the model. Motivated by this consideration, we are currently working on extending the architecture to a trisensory framework in which visual, tactile, and auditory pathways interact within the same network, allowing a more comprehensive account of cross-modal competition in naturalistic settings.

The sensitivity analyses further clarified the functional contribution of the interhemispheric cross-sensory inhibitory connections (Linter) on modality switching, accounting for the observed within and between-subject fluctuations in unisensory bilateral switch costs, in accordance with previous results in the audiovisual domain Cuppini et al. (2020). Recent studies in rodents and primates have identified GABAergic interneurons, such as somatostatin-positive and parvalbumin-positive cells, as key mediators of cross-modal suppression in early sensory cortices, suggesting that modality-specific inhibition may shape competition for neural resources. Whiskers stimulation has been found to suppress visually evoked activity in rodents, following the activation of long-range projections from the posterior primary somatosensory barrel cortex to fast-spiking inhibitory visual interneurons (Dinh et al., 2024; Weiler et al., 2024). EEG studies in humans (Cardini et al., 2011) also showed that visual input modulation of interneuronal circuits in the somatosensory cortex may mediate a suppression of locally evoked activity as a consequence of fingers stimulation, while fMRI-based works revealed further inhibitory neural modulation of visual cortical areas after tactile stimulation on the hand (Ide et al., 2016; Merabet et al., 2007).

Third, the network incorporates top-down interneurons that are driven by contralateral associative elements and suppress ipsilateral unisensory processing, realizing a interhemispheric top-down inhibitory mechanism, which is mediated by fast synapses and acts with a delay accounting for the callosal transfer time of neural information across hemispheres (Fendrich et al., 2004). Perturbing the effectiveness of top-down inhibitory connections (LInt) allowed us to identify critical variations in bisensory bilateral performance. In many instances, RTs to Sw trials were lower than those to Rp trials, i.e., the mean switch cost was negative. A plausible biological explanation is that repeating bilateral visuotactile stimulation maintains two highly salient multisensory events concurrently represented across the two hemispheres, a configuration that is more likely than unisensory stimulation to promote strong ipsilateral integration and, in turn, a stronger downstream recruitment of inhibitory control. Such inhibitory control, implemented both within hemisphere and via interhemispheric interactions, would delay contralateral processing on closely following events. Relative to a unisensory configuration, the repetition of bilateral crossmodal events may therefore increase interhemispheric competition for limited perceptual, attentional, and sensorimotor resources needed to generate a rapid response, thereby slowing response execution. In this context, switching from a unisensory event in one hemisphere to a crossmodal event in the other (i.e., a bilateral VT switch) may reduce the degree of simultaneous multisensory engagement and thus relax inhibitory demands, yielding faster responses than in repeat trials. More broadly, the finding aligns with literature on interhemispheric cross-sensory interactions, suggesting that corpus callosum-mediated inhibition can selectively modulate sensory salience depending on task demands (Berlucchi, 1995) and contribute to interhemispheric competition for access to attentional resources mediated by parietal and frontal structures (Mattingley et al., 1997; Hopfinger et al., 2000; Macaluso et al., 2000). Under this view, negative switch costs in bilateral VT conditions do not imply that switching is intrinsically beneficial; instead, they likely reflect a net competition penalty associated with maintaining two simultaneously salient multisensory representations across trials.

Furthermore, our network predictions offer valuable insight on how modality switching mechanisms may persist beyond immediate sensory processing, affecting even delayed perceptual decisions. Simulating a 4-alternative forced choice task (Wani et al., 2021), the network reproduced a drop in tactile detection accuracy when visual cues preceded tactile-only trials, suggesting that these offline effects may emerge from temporally extended cross-modal inhibition. Interestingly, the predicted accuracy performance resulted greater for the lower-intensity (Dim LED-associated) hand, as the cross-sensory inhibitory loop is less activated and the associated switching cost decreases. By contrast, empirical data showed a numerically opposite trend, i.e. higher accuracy for the Bright-associated hand, though this difference did not reach statistical significance. This divergence might reflect inter-individual variability in attentional weighting or perceptual salience of brighter stimuli, which should be explicitly modeled in a future architecture.

We will now present some limitations concerning the present work. First, the representation of neural populations in the model is considerably simplified, as pools of neurons that in the brain consist of distributed and heterogeneous assemblies were reduced to single computational units. Indeed, the primary aim here was not to reproduce the fine-grained encoding of sensory information but rather to investigate how dynamic interactions between modalities can give rise to specific behavioral patterns.

A second limitation concerns the empirical component of the study. While we are aware that the sample size is relatively small to provide a

solid validation to the network predictions, we believed that these preliminary results could be presented as an empirical complement to the presented network, providing an initial indication on whether the main qualitative patterns anticipated by the network can be observed in human behavior, without weakening the primarily computational emphasis of the work. The presented experimental results are not meant to provide a precise estimate of population-level effects, but the alignment between the measured patterns and the model's qualitative predictions suggests that it may be useful to report these data as a first step for evaluating the potential contribution of the feedback inhibitory loop, an element that has been explored less extensively in the visuotactile literature than in audiovisual integration. At the same time, extending these observations in a study with a larger sample will be important before drawing stronger conclusions about the neural underpinnings of bilateral switching.

Within this scope, the proposed network proposes a transparent and interpretable framework that brings together facilitative and competitive dynamics in bilateral visuohaptic processing, while remaining sufficiently modular to allow straightforward manipulation of units and connections to explore candidate functional alterations relevant to atypical multisensory behaviors, including those reported in children with neurodevelopmental disorders. For instance, Greenfield et al. (2015) demonstrated that individuals with autism spectrum disorder show atypical visuotactile integration likely due altered temporal binding that may result in a greater reliance on proprioceptive inputs. These findings also point to the need for a deeper characterization of visuotactile temporal dynamics in neurotypical populations, especially given that tactile exploration plays a critical role during early infancy for forming coherent multisensory representations.

Beyond its potential clinical relevance, the investigation of these mechanisms could also inform applications in robotics, where more natural and veridical integration of sensory cues would enhance the capacity of artificial agents to interact effectively with their environment. The temporal mechanisms instantiated in the BIVIP network may have direct relevance in the development of collaborative robots, also known as cobots, which are increasingly exploited in smart manufacturing applications to safely share workspaces and tasks with human operators (Addula & Tyagi, 2024). Cobots must continuously arbitrate between visual sensing, for instance in tasks of object localization, human pose/hand tracking, and tactile feedback, including contact detection or compliant control, hence our model offers a principled computational account of how facilitation and cross-sensory inhibition can stabilize these decisions when multisensory cues are synchronous, noisy, or temporally offset. Embedding BIVIP-like dynamics into a cobot's perception-to-action stack could support human-aware sensor fusion policies, amplifying congruent visuotactile evidence to accelerate confident actions, yet transiently suppressung competing modality streams to mitigate false triggers during rapid context changes (e.g., tool handovers, occlusions, or brief incidental contacts). In other words, because industrial settings often demand adaptable operation under changing layouts and product variants, a network that explicitly elucidate when and why to integrate versus compete across modalities may contribute to more agile, robust, and interpretable cobot behavior, complementing computer-vision-driven automation with temporally structured visuotactile control.

5. Conclusion

Altogether, these results provide novel evidence that the neural mechanisms underlying modality switching generalize across sensory pairs and spatial layouts. While the sample size presented along in complement to the modelling work was limited, it was meant to validate a network architecture previously tuned to extant behavioral data and inspired from previously proposed networks. Considering previous reports in the audiovisual domain, the consistency of results across sensory domains supports the notion that competitive and cooperative cross-modal dynamics represent domain-general neural computations. In summary, we suggest that the computations supporting modality switching extend to the visuotactile domain, operate across hemispheres, and persist beyond stimulus onset. Our neural network model links these behavioral effects to mechanistically interpretable neural circuits, highlighting how intersensory dynamics can shape perceptual decisions in the absence of direct stimulation.

CRediT authorship contribution statement

Eleonore Federica Di Rosa: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Aaisha Sheth:** Writing – review & editing, Visualization, Software, Data curation. **Jeffrey Min-In Yau:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Laura Astolfi:** Supervision, Resources, Project administration, Funding acquisition. **Cristiano Cuppini:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

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

This work was supported in part by funding from PRIN MUR 20207S3NB8 "ACT2 —Acting together: how motor styles shape action prediction and brain-to-brain connectivity in typical and autistic populations" and by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006)—A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022) and by AEGEUS - 101099210 - HORIZON - EIC - 2022 - PATHFINDEROPEN-01. Additional support for this work was provided by a grant from the National Science Foundation NSF (SBE) 2019959.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neunet.2026.108960.

References

Addula, S. R. (2024). Future of computer vision and industrial robotics in Smart manufacturing. In A. K. Tyagi, A. K. Tyagi, S. Tiwari, S. K. Arumugam, & A. K. Sharma (Eds.), *Artificial intelligence-enabled digital twin for smart manufacturing* (pp. 505–539). Wiley. https://doi.org/10.1002/9781394303601.ch22 (1st ed., pp.).

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex (New York, N.Y.: 1991), 12*(11), 1202–1212. https://doi.org/10.1093/cercor/12.11.1202

Anzola, G. P., Bertoloni, G., Buchtel, H. A., & Rizzolatti, G. (1977). Spatial compatibility and anatomical factors in simple and choice reaction time. *Neuropsychologia, 15*(2), 295–302. https://doi.org/10.1016/0028-3932(77)90038-0

Azañón, E., Stenner, M.-P., Cardini, F., & Haggard, P. (2015). Dynamic tuning of tactile localization to body posture. *Current Biology, 25*(4), 512–517. https://doi.org/10.1016/j.cub.2014.12.038

Badde, S., Heed, T., & Röder, B. (2016). Integration of anatomical and external response mappings explains crossing effects in tactile localization: A probabilistic modeling approach. *Psychonomic Bulletin & Review, 23*(2), 387–404. https://doi.org/10.3758/s13423-015-0918-0

Berlucchi, G. (1995). Corpus callosum and simple visuomotor integration. *Neuropsychologia, 33*(8), 923–936. https://doi.org/10.1016/0028-3932(95)00031-W

Berlucchi, G. (2006). Inhibition of return: A phenomenon in search of a mechanism and a better name. *Cognitive Neuropsychology, 23*(7), 1065–1074. https://doi.org/10.1080/02643290600588426

Boesch, E., Bellan, V., Moseley, G. L., & Stanton, T. R. (2016). The effect of bodily illusions on clinical pain: A systematic review and meta-analysis. *Pain, 157*(3), 516–529. https://doi.org/10.1097/j.pain.0000000000000423

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Research Bulletin, 75*(6), 753–760. https://doi.org/10.1016/j.brainresbull.2008.01.009

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience, 22*(11), 2886–2902. https://doi.org/10.1111/j.1460-9568.2005.04462.x

Cardini, F., Longo, M. R., & Haggard, P. (2011). Vision of the body modulates somatosensory intracortical inhibition. *Cerebral Cortex, 21*(9), 2014–2022. https://doi.org/10.1093/cercor/bhq267

Chivukula, S., Zhang, C. Y., Aflalo, T., Jafari, M., Pejsa, K., Pouratian, N., & Andersen, R. A. (2021). Neural encoding of actual and imagined touch within human posterior parietal cortex. *eLife, 10*, Article e61646. https://doi.org/10.7554/eLife.61646

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience, 16*(6), 1000–1009. https://doi.org/10.1162/0898929041502733

Craig, J. C., & Belser, A. N. (2006). The crossed-hands deficit in tactile temporal-order judgments: The effect of training. *Perception, 35*(11), 1561–1572. https://doi.org/10.1068/p5481

Craig, J. C., & Qian, X. (1997). Tactile pattern perception by two fingers: Temporal interference and response competition. *Perception & Psychophysics, 59*(2), 252–265. https://doi.org/10.3758/BF03211893

Crosse, M. J., Foxe, J. J., Tarrit, K., Freedman, E. G., & Molholm, S. (2022). Resolution of impaired multisensory processing in autism and the cost of switching sensory modality. *Communications Biology, 5*(1), 601. https://doi.org/10.1038/s42003-022-03519-1

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., & Ursino, M. (2014). A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage, 92*, 248–266. https://doi.org/10.1016/j.neuroimage.2014.02.001

Cuppini, C., Magosso, E., Monti, M., Ursino, M., & Yau, J. M. (2022). A neurocomputational analysis of visual bias on bimanual tactile spatial perception during a crossmodal exposure. *Frontiers in Neural Circuits, 16*, Article 933455. https://doi.org/10.3389/fncir.2022.933455

Cuppini, C., Ursino, M., Magosso, E., Crosse, M. J., Foxe, J. J., & Molholm, S. (2020). Cross-sensory inhibition or unisensory facilitation: A potential neural architecture of modality switch effects. *Journal of Mathematical Psychology, 99*, Article 102438. https://doi.org/10.1016/j.jmp.2020.102438

Dadarlat, M. C., O'Doherty, J. E., & Sabes, P. N. (2015). A learning-based approach to artificial sensory feedback leads to optimal integration. *Nature Neuroscience, 18*(1), 138–144. https://doi.org/10.1038/nn.3883

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

De Haan, A. M., Anema, H. A., & Dijkerman, H. C. (2012). Fingers crossed! an investigation of somatotopic representations using spatial directional judgements. *PLoS ONE, 7*(9), Article e45408. https://doi.org/10.1371/journal.pone.0045408

Di Rosa, E., Sheth, Aaisha, Y., Jeffrey, M., Astolfi, L., & Cuppini, C., (2025). *Unimanual* vs *bimanual intersensory switch in visuotactile detection: A multimodal analysis*. Convegno Nazionale di Bioingegneria. Patron Editore. https://hdl.handle.net/11585/1023673.

Diederich, A., & Colonius, H. (2007). Modeling spatial effects in visualtactile saccadic reaction time. *Perception & Psychophysics, 69*(1), 56–67. https://doi.org/10.3758/BF03194453

Dinh, T. N. A., Moon, H. S., & Kim, S.-G. (2024). Separation of bimodal fMRI responses in mouse somatosensory areas into V1 and non-V1 contributions. *Scientific Reports, 14*(1), 6302. https://doi.org/10.1038/s41598-024-56305-w

E. Donato, T. G. Thuruthel and E. Falotico, "Towards Interpretable Visuo-Tactile Predictive Models for Soft Robot Interactions," 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), Heidelberg, Germany, 2024,1567-1573, doi:10.1109/BioRob60516.2024.10719859.

Ehrsson, H. H. (2020). Multisensory processes in body ownership. *Multisensory Perception*, 179–200.

Fendrich, R., Hutsler, J. J., & Gazzaniga, M. S. (2004). Visual and tactile interhemispheric transfer compared with the method of Poffenberger. *Experimental Brain Research, 158*(1), 67–74. https://doi.org/10.1007/s00221-004-1873-6

Gentile, G., Petkova, V. I., & Ehrsson, H. H. (2011). Integration of visual and tactile signals from the hand in the Human brain: An fMRI study. *Journal of Neurophysiology, 105*(2), 910–922. https://doi.org/10.1152/jn.00840.2010

Greenfield, K., Ropar, D., Smith, A. D., Carey, M., & Newport, R. (2015). Visuo-tactile integration in autism: Atypical temporal binding may underlie greater reliance on proprioceptive information. *Molecular Autism, 6*(1), 51. https://doi.org/10.1186/s13229-015-0045-9

Hagmann, C. E., & Russo, N. (2016). Multisensory integration of redundant trisensory stimulation. *Attention, Perception, & Psychophysics, 78*(8), 2558–2568. https://doi.org/10.3758/s13414-016-1192-6

Hartcher-O'Brien, J., Gallace, A., Krings, B., Koppen, C., & Spence, C. (2008). When vision 'extinguishes' touch in neurologically-normal people: Extending the Colavita visual dominance effect. *Experimental Brain Research, 186*(4), 643–658. https://doi.org/10.1007/s00221-008-1272-5

Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research, 193*(2), 307–314. https://doi.org/10.1007/s00221-008-1626-z

Heed, T., & Azañón, E. (2014). Using time to investigate space: A review of tactile temporal order judgments as a window onto spatial processing in touch. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00076

Hillis, A. E., Chang, S., Heidler-Gary, J., Newhart, M., Kleinman, J. T., Davis, C., Barker, P. B., Aldrich, E., & Ken, L. (2006). Neural correlates of modality-specific spatial extinction. *Journal of Cognitive Neuroscience, 18*(11), 1889–1898. https://doi.org/10.1162/jocn.2006.18.11.1889

Hopfinger, JB, Buonocore, MH, & Mangun, GR (2000). The neural mechanisms of top-down attentional control. *Nat Neurosci, 3*(3), 284–291. https://doi.org/10.1038/72999

Holmes, NP, & Spence, C (2005). Multisensory integration: space, time and superadditivity. *Curr Biol, 15*(18), R762–4. https://doi.org/10.1016/j.cub.2005.08.058

Ide, M., Hidaka, S., Ikeda, H., & Wada, M. (2016). Neural mechanisms underlying touch-induced visual perceptual suppression: An fMRI study. *Scientific Reports, 6*(1), Article 37301. https://doi.org/10.1038/srep37301

Ioannides, A. A., Liu, L., Poghosyan, V., Saridis, G. A., Gjedde, A., Ptito, M., & Kupers, R. (2013). MEG reveals a fast pathway from somatosensory cortex to occipital areas via posterior parietal cortex in a blind subject. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00429

Jansen, B. H., & Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics, 73*(4), 357–366. https://doi.org/10.1007/BF00199471

Kandel. (2013). *Principles of neural science (5th ed)*. McGraw-Hill medical.

Kennett, S, Taylor-Clarke, M, & Haggard, P (2001). Noninformative vision improves the spatial resolution of touch in humans. *Curr Biol, 11*(15), 1188–1191. https://doi.org/10.1016/s0960-9822(01)00327-x

Lateiner, J. E., & Sainburg, R. L. (2003). Differential contributions of vision and proprioception to movement accuracy. *Experimental Brain Research, 151*(4), 446–454. https://doi.org/10.1007/s00221-003-1503-8

Lloyd, D. M., Mason, L., Brown, R. J., & Poliakoff, E. (2008). Development of a paradigm for measuring somatic disturbance in clinical populations with medically unexplained symptoms. *Journal of Psychosomatic Research, 64*(1), 21–24. https://doi.org/10.1016/j.jpsychores.2007.06.004

Lucan, J. N., Foxe, J. J., Gomez-Ramirez, M., Sathian, K., & Molholm, S. (2010). Tactile shape discrimination recruits human lateral occipital complex during early perceptual processing. *Human Brain Mapping, 31*(11), 1813–1821. https://doi.org/10.1002/hbm.20983

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences, 28*(5), 264–271. https://doi.org/10.1016/j.tins.2005.03.008

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science (New York, N.Y.), 289*(5482), 1206–1208. https://doi.org/10.1126/science.289.5482.1206

Macaluso E, Frith CD, Driver J. Modulation of human visual cortex by crossmodal spatial attention. Science. 2000; 289 (5482) : 1206-8. doi:10.1126/science.289.5482.1206.

Magosso, E., Serino, A., Di Pellegrino, G., & Ursino, M. (2010). Crossmodal links between vision and touch in spatial attention: A computational modelling study. *Computational Intelligence and Neuroscience, 2010*, 1–13. https://doi.org/10.1155/2010/304941

Marzi, C. A., Bisiacchi, P., & Nicoletti, R. (1991). Is interhemispheric transfer of visuomotor information asymmetric? Evidence from a meta-analysis. *Neuropsychologia, 29*(12), 1163–1177. https://doi.org/10.1016/0028-3932(91)90031-3

Mattingley, J. B., Driver, J., Beschin, N., & Robertson, I. H. (1997). Attentional competition between modalities: Extinction between touch and vision after right hemisphere damage. *Neuropsychologia, 35*(6), 867–880. https://doi.org/10.1016/S0028-3932(97)00008-0

Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology, 68*(4), 1332–1344. https://doi.org/10.1152/jn.1992.68.4.1332

Merabet, L. B., Hamilton, R., Schlaug, G., Swisher, J. D., Kiriakopoulos, E. T., Pitskel, N. B., Kauffman, T., & Pascual-Leone, A. (2008). Rapid and reversible recruitment of early visual cortex for touch. *PLoS ONE, 3*(8), e3046. https://doi.org/10.1371/journal.pone.0003046

Merabet, L. B., Swisher, J. D., McMains, S. A., Halko, M. A., Amedi, A., Pascual-Leone, A., & Somers, D. C. (2007). Combined activation and deactivation of visual cortex during tactile sensory processing. *Journal of Neurophysiology, 97*(2), 1633–1641. https://doi.org/10.1152/jn.00806.2006

Mirams, L., Poliakoff, E., & Lloyd, D. M. (2017). Spatial limits of visuotactile interactions in the presence and absence of tactile stimulation. *Experimental Brain Research, 235*(9), 2591–2600. https://doi.org/10.1007/s00221-017-4998-0

Nicoletti, R., Umiltà, C., & Ladavas, E. (1984). Compatibility due to the coding of the relative position of the effectors. *Acta Psychologica, 57*(2), 133–143. https://doi.org/10.1016/0001-6918(84)90039-8

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology, 2*(3), 211–228. https://doi.org/10.1080/02643298508252866

Ramos-Estebanez, C., Merabet, L. B., Machii, K., Fregni, F., Thut, G., Wagner, T. A., Romei, V., Amedi, A., & Pascual-Leone, A. (2007). Visual phosphene perception modulated by subthreshold crossmodal sensory stimulation. *The Journal of Neuroscience, 27*(15), 4178–4181. https://doi.org/10.1523/JNEUROSCI.5468-06.2007

Reed, J. L., Qi, H.-X., Zhou, Z., Bernard, M. R., Burish, M. J., Bonds, A. B., & Kaas, J. H. (2010). Response properties of neurons in primary somatosensory cortex of owl monkeys reflect widespread spatiotemporal integration. *Journal of Neurophysiology, 103*(4), 2139–2157. https://doi.org/10.1152/jn.00709.2009

Risso, G., Valle, G., Iberite, F., Strauss, I., Stieglitz, T., Controzzi, M., Clemente, F., Granata, G., Rossini, P. M., Micera, S., & Baud-Bovy, G. (2019). Optimal integration of intraneural somatosensory feedback with visual information: A single-case study. *Scientific Reports, 9*(1), 7916. https://doi.org/10.1038/s41598-019-43815-1

Rizzolatti, G, Fadiga, L, Fogassi, L, & Gallese, V (1997). *The space around us. Science, 277* (5323), 190–191. https://doi.org/10.1126/science.277.5323.190

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50*(1–2), 19–26. https://doi.org/10.1016/S0167-8760(03)00121-1

Rognini, G., Petrini, F. M., Raspopovic, S., Valle, G., Granata, G., Strauss, I., Solcà, M., Bello-Ruiz, J., Herbelin, B., Mange, R., D'Anna, E., Di Iorio, R., Di Pino, G., Andreu, D., Guiraud, D., Stieglitz, T., Rossini, P. M., Serino, A., Micera, S., & Blanke, O. (2019). Multisensory bionic limb to achieve prosthesis embodiment and reduce distorted phantom limb perceptions. *Journal of Neurology, Neurosurgery & Psychiatry, 90*(7), 833–836. https://doi.org/10.1136/jnnp-2018-318570

Samad, M., & Shams, L. (2016). Visual-somatotopic interactions in spatial perception. *Neuroreport, 27*(3), 180–185. https://doi.org/10.1097/WNR.0000000000000521

Sarri, M., Blankenburg, F., & Driver, J. (2006). Neural correlates of crossmodal visual-tactile extinction and of tactile awareness revealed by fMRI in a right-hemisphere stroke patient. *Neuropsychologia, 44*(12), 2398–2410. https://doi.org/10.1016/j.neuropsychologia.2006.04.032

Schürmann, M., Kolev, V., Menzel, K., & Yordanova, J. (2002). Spatial coincidence modulates interaction between visual and somatosensory evoked potentials. *NeuroReport*, (6), 13. https://journals.lww.com/neuroreport/fulltext/2002/05070/spatial_coincidence_modulates_interaction_between.9.aspx.

Shaw, L. H., Freedman, E. G., Crosse, M. J., Nicholas, E., Chen, A. M., Braiman, M. S., Molholm, S., & Foxe, J. J. (2020). Operating in a multisensory context: Assessing the interplay between multisensory reaction time facilitation and inter-sensory task-switching effects. *Neuroscience, 436*, 122–135. https://doi.org/10.1016/j.neuroscience.2020.04.013

Sherrick, C. E. (1964). Effects of double simultaneous stimulation of the skin. *The American Journal of Psychology, 77*(1), 42. https://doi.org/10.2307/1419270

Soto-Faraco, S., Ronald, A., & Spence, C. (2004). Tactile selective attention and body posture: Assessing the multisensory contributions of vision and proprioception. *Perception & Psychophysics, 66*(7), 1077–1094. https://doi.org/10.3758/BF03196837

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63*(2), 330–336. https://doi.org/10.3758/bf03194473

Spence C, Pavani F, Maravita A, Holmes N. Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: evidence from the crossmodal congruency task. J Physiol Paris. 2004; 98 (1-3):171-89. doi:10.1016/j.jphysparis.2004.03.008. PMID: 15477031.

Tamè, L., Farnè, A., & Pavani, F. (2011). Spatial coding of touch at the fingers: Insights from double simultaneous stimulation within and between hands. *Neuroscience Letters, 487*(1), 78–82. https://doi.org/10.1016/j.neulet.2010.09.078

Tamè, L., & Longo, M. R. (2015). Inter-hemispheric integration of tactile-motor responses across body parts. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00345

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network: Computation in Neural Systems, 4*(3), 259–284. https://doi.org/10.1088/0954-898X_4_3_002

Van Der Groen, O., Van Der Burg, E., Lunghi, C., & Alais, D. (2013). Touch influences visual perception with a tight orientation-tuning. *PLoS ONE, 8*(11), Article e79558. https://doi.org/10.1371/journal.pone.0079558

Vanneau, T., Quiquempoix, M., Foxe, J. J., & Molholm, S. (2025). Neural mechanisms of intersensory switching: Evidence for delayed sensory processing and increased cognitive demands. *NeuroImage, 309*, Article 121089. https://doi.org/10.1016/j.neuroimage.2025.121089

Verrillo, R. T., Gescheider, G. A., Calman, B. G., & Van Doren, C. L. (1983). Vibrotactile masking: Effects of oneand two-site stimulation. *Perception & Psychophysics, 33*(4), 379–387. https://doi.org/10.3758/BF03205886

Wani, Y. R., Convento, S., & Yau, J. M. (2021). Vision automatically exerts online and offline influences on bimanual tactile spatial perception. *Journal of Mathematical Psychology, 100*, Article 102480. https://doi.org/10.1016/j.jmp.2020.102480

Weiler, S., Rahmati, V., Isstas, M., Wutke, J., Stark, A. W., Franke, C., Graf, J., Geis, C., Witte, O. W., Hübener, M., Bolz, J., Margrie, T. W., Holthoff, K., & Teichert, M. (2024). A primary sensory cortical interareal feedforward inhibitory circuit for tacto-visual integration. *Nature Communications, 15*(1), 3081. https://doi.org/10.1038/s41467-024-47459-2

Yamamoto, S., & Kitazawa, S. (2001). Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience, 4*(7), 759–765. https://doi.org/10.1038/89559