Journal of Mathematical Psychology 99 (2020) 102438

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Cross-sensory inhibition or unisensory facilitation: A potential neural architecture of modality switch effects

Cristiano Cuppini [a,*], Mauro Ursino [a], Elisa Magosso [a], Michael J. Crosse [b,c], John J. Foxe [b,d], Sophie Molholm [b,d]

[a] Department of Electrical, Electronic, and Information Engineering Guglielmo Marconi, University of Bologna, Bologna, Italy
[b] The Sheryl and Daniel R. Tishman Cognitive Neurophysiology Laboratory, Albert Einstein College of Medicine, Bronx, NY, United States of America
[c] X, the moonshot factory, Mountain View, CA, United States of America
[d] Ernest J. Del Monte Institute for Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, NY, United States of America

ARTICLE INFO

Article history:
Received 8 April 2020
Received in revised form 22 July 2020
Accepted 16 August 2020
Available online xxxx

Keywords:
Inhibitory mechanisms
Integrative mechanisms
Neural network
Sensory processing dynamics

ABSTRACT

In a simple reaction time task in which auditory and visual stimuli are presented in random sequence alone (A or V) or together (AV), there is a so-called reaction time (RT) cost on trials in which sensory modality switches (A→V) compared to when it repeats (A→A). This is always true for unisensory trials, whereas RTs to AV stimuli preceded by unisensory stimuli are statistically comparable with the Repeat condition (AV→AV). Neural facilitation for Repeat trials or neural inhibition for Switch trials could both account for these effects. Here we used a neural network model (Multisensory Integration with Crossed Inhibitory Dynamics (MICID) model) to test the ability of these two distinct mechanisms, inhibition and facilitation, to produce the specific patterns of behavior that we see experimentally, modeling switch and repeat trials as well as the influence of the interval between the present and the previous trial. The model results are consistent with an inhibitory account in which there is competition between the different sensory modalities, instead of a facilitation account in which the preceding stimulus sensitizes the neural system to its particular sensory modality. Moreover, the model shows that multisensory integration can explain the results in case of multisensory stimuli, where the preceding stimulus has little effect. This is due to faster dynamics for multisensory facilitation compared to cross-sensory inhibition. These findings link the cognitive framework delineated by the empirical results to a plausible neural implementation.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Behavioral data from Crosse, Foxe, and Molholm (2019) and Shaw et al. (2020) show that, in a simple reaction time task in which auditory and visual stimuli are presented, there is a behavioral cost when the modality of the stimulus switches (the Switch Condition, "Sw") compared to when it repeats (Repeat Condition, "Rp"), such that the Sw condition leads to comparatively longer RTs. Known as the modality switch effect (MSE), this is always true in the case of unisensory stimuli (A or V). In contrast, in the case of multisensory inputs, the RT to an AV stimulus, when preceded by an A or V stimulus, is statistically comparable with the Rp condition, where the preceding input is also multisensory (AV).

Two neural architectures may account for these patterns of effects. In the first (named *unisensory facilitation*) the preceding stimulus activates and biases the neural system to that particular sensory modality. In this way it exerts a facilitatory effect on the processing of the following stimulus when of the same sensory modality, leading to speeded RTs. This mechanism would not directly impact processing of a stimulus of a different sensory modality, and thus the Sw condition would not be affected. The second possibility (named *cross-sensory inhibition*) involves an opposite effect in which an initial sensory input inhibits the opposing (non-stimulated) sensory modality, so that the processing of a second stimulus of a different modality is penalized, resulting in the brain reacting slower to a Sw condition compared to a Rp one. This hypothesis assumes that competition among sensory modalities in the brain is the main factor accounting for these behavioral patterns. These two possibilities require different neural architectures and predict different results in the case of specific input configurations.

In particular, the time course of these two distinct mechanisms would lead to different patterns of results for short versus long inter-stimulus intervals (ISIs). For example, if two stimuli of the same sensory modality are presented in succession (Rp condition), according to a unisensory facilitation account, RTs to the

* Correspondence to: Viale Risorgimento 2, 40136 Bologna, Italy.
E-mail address: cristiano.cuppini@unibo.it (C. Cuppini).

https://doi.org/10.1016/j.jmp.2020.102438
0022-2496/© 2020 Elsevier Inc. All rights reserved.

second stimulus would be faster for shorter ISIs (when the facilitatory mechanism is still more active) versus longer ISIs (when the effect of the facilitatory mechanism is vanishing). Conversely, in the presence of a cross-modal inhibition, the effect would be apparent in Sw trials only, and RT would be increased in case of shorter ISI (due to a stronger inhibition) compared with longer ISI (when the mechanism is vanishing). These two alternative hypotheses may be tested by analyzing the RTs obtained by Crosse, Foxe, and Molholm (2019) for Rp and Sw trials separately for short versus long ISIs.

However, an additional mechanism must be considered, besides the two possible mechanisms mentioned above. Indeed, it is well known that cross-modal stimuli occurring in spatial and temporal proximity produce a reinforced response, a phenomenon usually named *multisensory integration* (MSI).

Several studies analyzed the effect that temporal misalignment of stimuli of different sensory modality may have on facilitative multisensory integration (MSI) effects, under different experimental conditions and at different levels of observation, from the neural (Meredith, Nemitz, & Stein, 1987; Miller et al., 2015; Stein & Meredith, 1993) to the behavioral perspective, (e.g., Bell et al., 2005, 2006; Colonius & Diederich, 2004; Lewald, Ehrenstein, & Guski, 2001; Lewald & Guski, 2003; Mégevand et al., 2013; Meredith, 2002; Musacchia & Schroeder, 2009; Navarra et al., 2005; Parise et al., 2013; Romei et al., 2007; Rowland & Stein, 2007; Rowland et al., 2007; Spence & Squire, 2003; Stevenson & Wallace, 2013; Wallace et al., 2004; van Wassenhove, Grant, & Poeppel, 2007). These data suggest that two stimuli of different sensory modalities produce a facilitatory effect, if spatially and temporally coincident; however, in case of temporal misalignment, this facilitation decays after hundreds of milliseconds (the Temporal Window of Integration, TWI, can vary from 40 to 400 ms, depending on context). Hence, the temporal dynamics describing this multisensory process are characterized by a short time course. Conversely, it is reasonable to assume that the inhibition among sensory modalities, hypothesized above as a possible mechanism affecting the Sw trials, lasts longer than the TWI. This assumption is consistent with the idea that two stimuli of different sensory modality must be integrated if occurring in spatial and temporal proximity, but kept segregated if occurring at large spatial and/or temporal disparity.

Indeed, to be consistent with the same multisensory integrative rules observed in the spatial domain, that two stimuli of different sensory modality are integrated, as long as the reciprocal distance is not too large, then, once the spatial disparity becomes too large, they reciprocally inhibit, a similar behavior can be expected to occur also in the temporal domain, but on a different time scale.

In the following, the results by Crosse, Foxe, and Molholm (2019) are analyzed first, to discriminate between the neural mechanisms delineated above. Then, a neural architecture is proposed to provide an interpretation for these results (Crosse, Foxe, & Molholm, 2019).

1.1. Data analysis from Crosse, Foxe, and Molholm' experiment (2019)

To discriminate between these two architectures and neural mechanisms delineated above, the first step is to analyze the experimental results of Crosse, Foxe, and Molholm (2019) in the case of short (< 1500 ms) and long (> 2500 ms) ISIs. If RTs in Sw conditions are slower for short ISIs than for long ISIs, this would support competition among sensory modalities as the main mechanism underlying the MSE observed in these tasks. Conversely, shorter RTs in repeat conditions for short ISIs would support a facilitation mechanism.

To better understand the behavioral results analyzed in this section, we briefly summarize the experimental procedures of Crosse, Foxe, and Molholm (2019).

In this experiment, researchers analyzed data from participants ranging in age from 6 to 36 years and with neurotypical development or a diagnosis of Autism Spectrum Disorder. Here we consider the data from the neurotypical adults (NT, age range: 18–40 years, $n = 70$).

The stimulus materials were identical to those described in Brandwein et al. (2011). Visual (V) stimuli consisted of a red disc (diameter: 3.2 cm; duration: 60 ms), located 0.4 cm above a central fixation crosshair on a black background. Auditory (A) stimuli consisted of a 1 kHz pure tone, sampled at 44.1 kHz (duration: 60 ms; rise/fall time: 5 ms). Audiovisual (AV) stimuli consisted of the combined simultaneous pairing of the auditory and visual stimuli described above.

Participants performed a speeded detection task on a computer and were seated 122 cm from the visual display in a dimly-lit, sound-attenuated booth. To reduce predictability, the stimuli were presented in a completely randomized order with equal probability and the ISI was randomly jittered between 1000–3000 ms according to a uniform, square-wave distribution. Stimulus presentation was controlled using Presentation® software (Neurobehavioral Systems, Inc., Berkeley, CA). Auditory stimuli were delivered binaurally at an intensity of 75 dB SPL via a single, centrally-located loudspeaker (JBL Duet Speaker System, Harman Multimedia). Visual stimuli were presented at a resolution of 1280 × 1024 pixels on a 17 inch Flat Panel LCD monitor (Dell Ultrasharp 1704FTP). The auditory and visual stimuli were presented in close spatial proximity, with the speaker placed atop the monitor and aligned vertically to the visual stimulus. Participants were instructed to press a button with their right thumb as soon as they perceived any of the three stimuli. Latencies of stimulus onsets and button presses were acquired and stored digitally at a sampling rate of 512 Hz. Stimuli were presented in blocks of ∼100 trials and participants typically completed 6–10 blocks in total.

Response times were measured relative to the onset time of the preceding stimulus and analyzed separately for each participant. An outlier correction procedure was performed before the main RT analyses. First, RTs that did not fall within 100–2000 ms post-stimulus were removed. On average, fast outliers (< 100 ms, considered anticipatory responses) made up 0.7% (±0.9) of trials and slow outliers (> 2000 ms, considered misses) made up 0.4% (±0.6) of trials. Second, RTs outside the middle 95th percentile (2.5–97.5) of their respective conditions were removed. Stimuli were categorized as either switch or repeat trials based on the modality of the preceding stimulus (repeat trials: AV→AV, A→A, V→V; switch trials: V→AV, A→AV, A→V, V→A). Trials AV→A and AV→V were excluded from the analysis as they were considered neither switches nor repeats.

Since the aim of the neurocomputational model described here is to suggest the more plausible neural architecture responsible for the behaviors reported in Crosse, Foxe, and Molholm (2019), we implement a model simulating the "NT adult" behavior. So in this section, we analyze data collected from the NT adult group only (18–40 years, $n = 70$).

We performed a 2-way ANOVA with factors of condition (switch vs repeat) and ISI duration (short vs long) on the mean RTs of each NT adult subject, computed over the repetitions for each analyzed condition, as detailed later. On average 25 responses per participant were represented in each condition for each individual subject's average (except the multisensory switch trials which had double the number of possible trials), and this did not differ meaningfully for the short or long ISIs (at 24 and 26 each). Approximately 25% of all possible responses

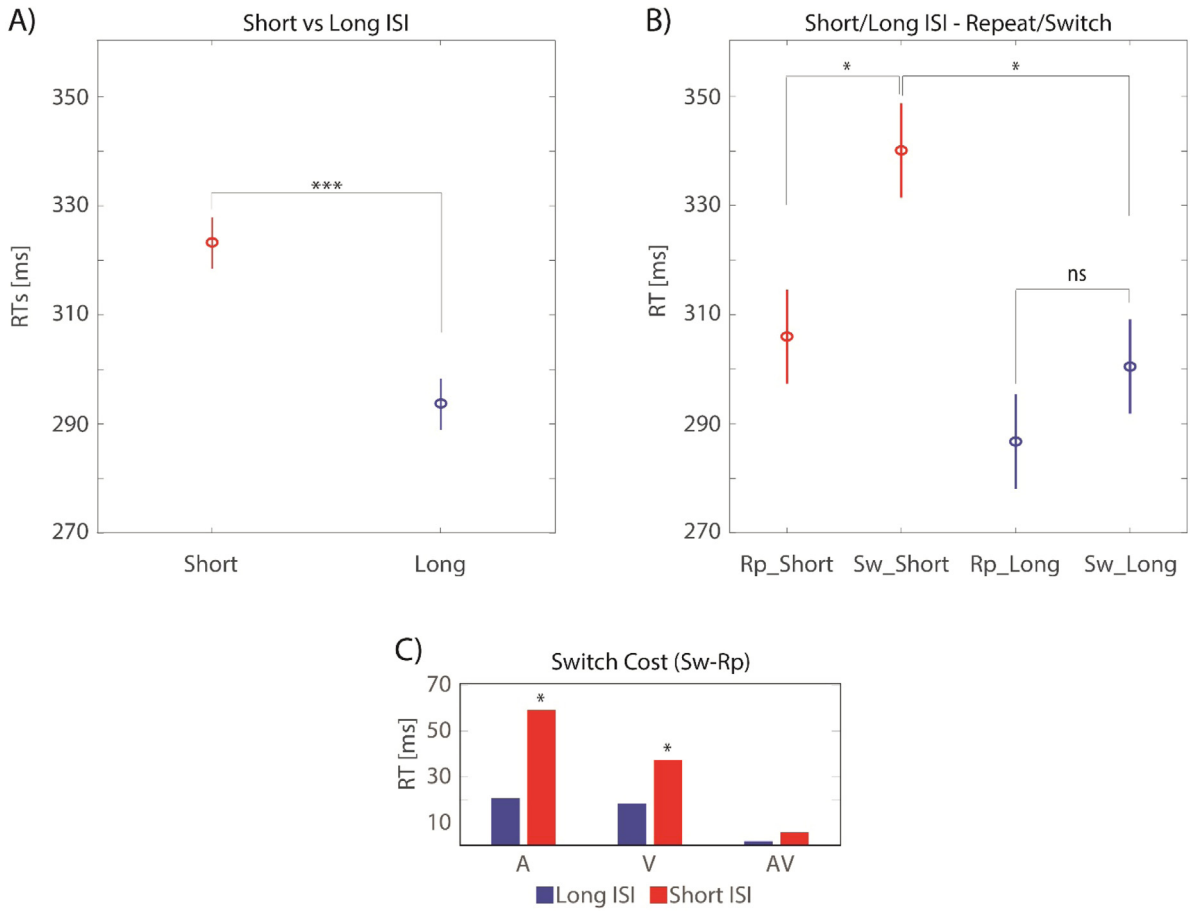

Fig. 1. Statistical Analysis of the effect of ISIs on the RTs. (A) Reaction times (collapsed across A, V, AV conditions) in case of short ISIs (< 1500 ms) are longer than RTs obtained with long ISIs (> 2500 ms). (B) Investigating the effect of ISIs on the Switch/Repeat comparison, it is evident that RTs are significantly different only in case of fast ISIs (red lines). Vice versa, for stimuli presented with ISIs > 2500 ms (blue lines) the effect of switching sensory modality is no longer significant. Moreover, the switch cost (RTs in Switch condition-RTs in Repeat condition) is significantly reduced in case of long ISIs: RTs in Sw condition are significantly faster in case of long ISIs than in case of short ISIs. (C) A further analysis on the Switch Cost reveals that Repeat vs Switch RTs are significantly different only for unisensory auditory and visual conditions with short ISIs, but not in case of AV stimuli.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were represented in each of the short and long ISI bins. As shown in Fig. 1, this analysis showed significant differences for both, condition ($F_{1/483} = 25.28$, $p < 0.0001$) and ISI duration ($F_{1/483} = 38.26$, $p < 0.0001$).

As can be seen in Fig. 1A, RTs were strongly affected by ISI such that responses were significantly faster for longer (> 2500 ms) compared to shorter ISIs (< 1500 ms). More critically, an interaction between ISI and repeat versus switch condition lends support for a competitive mechanism between the sensory systems underlying the switch effects. In the case of competition, short time intervals should lead to greater competition between the sensory modalities and thus larger switch effects when compared to longer ISIs. As shown in Fig. 1B, this is true: for the shorter ISIs, the switch condition presents the longest RT, across all conditions, and the difference between switch and repeat is significant and larger than for the longer ISI condition. In case of longer ISIs, there is an overall RT advantage and the difference between switch and repeat is no longer statistically significant.

These results strongly suggest the presence of a competition between the two sensory modalities that decreases with increasing temporal distance between the stimuli and thus motivate us to develop a model in which competition between modalities serves as the primary cause of the observed sensory modality switch effects. Further support for this hypothesis is given by the comparison of Switch Costs when evaluated separately by stimulus conditions (Fig. 1C), as a function of long vs short ISIs. From this analysis, it is clear that the main inhibitory effect happens in case of unisensory switch conditions, for both modalities, for the short but not long ISIs, while it is not statistically significant in case of multisensory stimuli. This model is described in the next section, followed by a comparison of the model results with the behavioral data from Crosse, Foxe, and Molholm (2019).

2. Method

2.1. Basal model: qualitative description

The model has been realized to simulate a behavioral task where subjects were required to respond to auditory and visual stimuli, alone or combined, presented in a random sequence.

The model consists of 3 layers. The first layer, representing auditory and visual areas (A and V in Fig. 2) is the "*input layer*". It receives external stimuli of the corresponding sensory modalities and provides the first sensory processing step. External stimuli are excitatory inputs with an assigned efficacy, chosen to elicit strong activity in the input regions, and a duration of 60 ms, presented at a rate of every 1 to 3 s (boxcar function; millisecond steps). The onset, duration, and presentation rate (inter-stimulus interval: ISI) of these stimuli were chosen to mimic the experimental setup of Crosse, Foxe, and Molholm (2019), whose data

Table 1
Parameters value.

Neurons			
$\theta = 25$		$s = 0.3$	$\tau = 3$ ms
Inputs			
$G_i^r = 75$; $i = e, c, I\_ex, I\_in, m$	$\tau_i^r = 15$ ms; $i = c, I\_ex, I\_in, m$	$\tau_e^a = 15$ ms	$\tau_e^v = 25$ ms
$I_0^a = [1.09 - 1.21]$	$T^a = 60$ ms	$I_0^v = [1.6 - 1.9]$	$T^v = 60$ ms
Inhibition			
$G_l^r = 750; r = a, v$		$\tau_l^r = 180$ ms; $r = a, v$	
Synapses			
$W = 0.2$	$\Delta t = 16$ ms	$L = 0.1$	$WI = 2$
$LI = 3$		$Wm = 3$	$\Delta t^m = 100$ ms

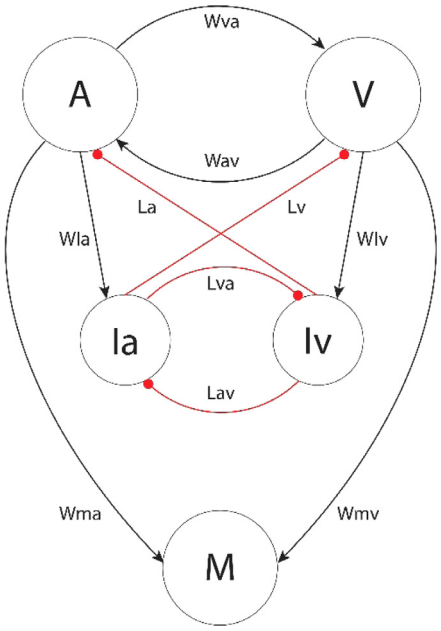

Fig. 2. Model Structure: A and V represent the auditory and visual regions, responsible for the first sensory processing, and implementing the input layer. They exchange direct excitatory synapses, representing a cross-modal interaction. M is the multisensory/motor output region. It is responsible for generating the simulated reaction times (RTs) to the external stimuli. *Ia* and *Iv* are unisensory inhibitory areas, excited by the input layer and implementing a competitive mechanism between modalities via winner-takes-all dynamics. Black lines represent excitatory connections; red lines represent inhibitory synapses.

are compared with our results. Moreover, to reproduce experimental variability, the external input (I_0^v, and I_0^a,) is randomly taken from a uniform distribution (see Table 1).

Besides these external stimuli, neural elements in these two areas receive two additional inputs: A cross-modal excitatory component from the other sensory region, and inhibitory feedback from a competitive layer:

- Auditory and visual regions are directly interconnected through excitatory cross-modal synapses, *Wav* and *Wva*, characterized by fast dynamics. So, in case of multisensory presentation, these cross-modal synapses generate an excitatory component, which increases the overall excitation of each region. This increase is present only if the region is already stimulated by an external input of the corresponding sensory modality. In fact, the strength of these synapses, W, has been set so that, in unisensory conditions, activity in one region is not able to activate the other sensory region. However, this strength is strong enough to induce a significant increase in the activity of the other sensory input area in case of a multisensory stimulation. This is in line with findings from animal studies in which cross-sensory inputs modulate excitation levels of another sensory region but do not generate a full response by themselves. Presented in conjunction with the primary stimulus however, they result in an enhanced response (Bizley & King, 2008, 2009; Ghazanfar & Schroeder, 2006; Kayser, Petkov, & Logothetis, 2008; Meredith & Allman, 2015; Yu et al., 2013).
- The inhibitory contribution from the competitive layer implements cross-sensory competition between different sensory modalities when the stimuli are presented sequentially. Modality specific neurons in this layer produce long-lasting inhibition of the region processing external inputs of the other sensory modality, through inhibitory feedback synapses. These inhibitory synapses are characterized by slow dynamics, responsible for the long-lasting inhibitory effect.

Excitatory synapses (*WIa* and *WIv*) project from the input regions to modality specific inhibitory areas (Ia and Iv, see Fig. 2), and long-range excitatory connections (*Wma* and *Wmv* in Fig. 2) to a read-out layer, simulating a multisensory/motor area (M region in Fig. 2).

The middle layer, named "*competitive layer*", implements a competitive mechanism by means of reciprocal inhibitory synapses, *Lav* and *Lva*, producing a winner-takes-all dynamic. With this structure, the "winning" sensory modality can exert the inhibitory effect on the other modality, as previously described, through feedback inhibitory synapses (*La* and *Lv* in Fig. 2), characterized by slow dynamics. Even if experimental evidence supports the idea of competition among sensory modalities, the neural underpinnings of this cross-modal inhibition are not yet clear. This competitive layer can be implemented through higher order regions, for example located in the medial Prefrontal Cortex (mPFC) or the Posterior Cingulate Cortex (PCC). Huang et al. (2015) suggested that these regions may be involved in the competition between A and V sensory modalities, in a simple RT experiment; Hairston et al. (2008) reached a similar conclusion in the case of an auditory temporal order judgment (TOJ) task. However, the competitive layer can also represent a direct influence between sensory regions: anatomical studies in animal models have shown projections between core visual and auditory regions and associative areas (Cappe & Barone, 2005; Clavagnier, Falchier, & Kennedy, 2004), and revealed direct modulatory functional connections among sensory regions (Bizley & King, 2008, 2009; Meredith & Allman, 2015; Yu et al., 2013).

Finally, the third layer, the multisensory output area, is used to mimic the behavioral responses of subjects to external stimuli: the elicited activity of this output region is compared with a fixed threshold (30% of the maximum neurons' activity) to evaluate the simulated RTs to the external stimuli.

To sum up, the architecture of the model implements two multisensory effects, one facilitatory and the other inhibitory. (1) The input layer and the multisensory area, with their specific synaptic architecture (i.e., the reciprocal excitatory connections, and the converging feedforward synapses), characterized by fast dynamics, provide the neural substrates for the multisensory processes performed in response to an external input, and, in case of audiovisual stimulation, implement multisensory facilitation. (2) The feedback projections to the input layer from the competitive layer, described by slower dynamics, realize competition between the sensory modalities.

So this network is characterized by multisensory facilitation with fast dynamics and a cross-modal competition with a long-lasting slow dynamics. In this way, not only can the model simulate the temporal profile of sensory processing in the brain, in case of unisensory and multisensory stimulations, but it is also able to explain how a first stimulus can affect the ability of the brain to process and react to a subsequent input of a different sensory modality. Due to the mechanisms included in this architecture, we named it as the "Multisensory Integration with Crossed Inhibitory Dynamics" (MICID) model.

In the model, each region is simulated with a single neural element. This choice is made for simplicity; given the experimental set-up of Crosse, Foxe, and Molholm (2019), we do not require either multiple units sensitive to a different spatial position in each sensory area as implemented in our previous neurocomputational spatial models (see, Cuppini, Magosso, & Ursino, 2011; Cuppini, Stein, & Rowland, 2018; Cuppini et al., 2011; Cuppini et al., 2012, 2014, 2017a; Magosso, Cuppini, & Ursino, 2012; Ursino, Cuppini, & Magosso, 2017; Ursino et al., 2017; Ursino et al., 2019) nor do we need multiple sensory regions sensitive to different input features, as necessary to realize semantic memory models (Cuppini, Magosso, & Ursino, 2009; Ursino, Cuppini, & Magosso, 2010, 2011; Ursino, Magosso, & Cuppini, 2009b; Ursino et al., 2018).

2.2. The basal model: mathematical description

For simplicity, each region of the model is described by a single neural element. Every element has been described by means of a first order differential equation, which simulates the integrative properties of the cellular membrane, and a steady-state sigmoidal relationship that simulates the presence of a lower threshold and an upper saturation for neural activation. The saturation value is set at 1, i.e., all outputs are normalized to the maximum. The term "activity" is used to denote the output of each area.

In the following, each element will be denoted with a superscript, r, referred to a specific region of the model ($r = a$, v, m, ia, or iv, where a refers to the auditory input area, v to the visual input area, m to the multisensory/motor output region, ia and iv to the inhibitory auditory and visual neurons, respectively). $u(t)$ and $y(t)$ are used to represent the net input and output of a given neural element at time t, respectively. Thus, $y^r(t)$ represents the output of the neural element simulating the region r, described by the following differential equation:

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

where τ is the time constant and $F(u)$ represents the sigmoidal relationship:

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

s and θ are parameters which establish the slope and the central position of the sigmoidal relationship, respectively.

For the sake of simplicity, in this work all the neural elements are described by using the same parameters and the same time constant.

The net input that reaches a specific neural element (i.e., the quantity $u^r(t)$ in Eq. (1)) depends on the region it belongs to.

Input areas — Elements in these regions process separately auditory and visual external stimuli ($r = a$, v). Their net input is the result of three components.

The first, the "external" component, is the unisensory input $e^r(t)$, coming from the external world. The second, the "cross-modal" component, is the input, $c^r(t)$, from the area processing the other sensory modality, transmitted to the target neuron through excitatory synapses. The last, the "inhibitory" component, is the contribution of the feedback inhibitory synapses from the interneuron activated by the other sensory modality, $I^r(t)$.

The external input is characterized by its effectiveness I_0^r, and its duration T^r. Assuming a stimulus of sensory modality r ($r = a$ or v):

$$e^r(t) = \begin{cases} I_0^r, 0 \le t \le T^r \\ 0, t > T^r \end{cases} \tag{3}$$

The cross-modal input, $c^r(t)$, is obtained assuming that each element receives an excitation from the region processing the other modality defined as:

$$\begin{aligned} c^a(t) &= W_{av} \cdot y^v(t - \Delta t) \\ c^v(t) &= W_{va} \cdot y^a(t - \Delta t) \end{aligned} \tag{4}$$

Where W_{va}, W_{av} are the weights of this reciprocal excitation. In the model they are the same for both regions ($W_{va} = W_{av} = W$), and Δt is a delay which represents the latency with which cross-modal inputs are exchanged between the two regions.

Finally, the inhibitory input, $I^r(t)$, is the effect of the interneuron excited by the other input sensory region that interacts with the target element through inhibitory synapses. Thus, the input that a unisensory element receives from the interneuron of the other modality is defined as:

$$\begin{aligned} I^a(t) &= L_a \cdot y^{iv}(t) \\ I^v(t) &= L_v \cdot y^{ia}(t) \end{aligned} \tag{5}$$

where $y^{ia}(t)$ and $y^{iv}(t)$ are the activities of presynaptic auditory and visual interneurons respectively, and L_a, L_v are the strengths of the inhibitory synapses. These synapses are symmetrical ($L_a = L_v = L$). In the model, we do not incorporate a delay for the cross-sensory inhibition, because the dynamics of the inhibitory effect has been chosen much longer than every other mechanism of the network (see below *Dynamics of each input component*), so the effect of a delay is already included in the time constant chosen for these projections.

Inhibitory interneurons — elements in these two regions ($r = ia, iv$) are excited respectively by the auditory and visual input areas, and they exchange inhibitory projections, implementing a "winner-takes-all" (WTA) mechanism. Their net input is the result of an excitatory stimulus, $I\_ex^r(t)$, coming from the corresponding unisensory input region through excitatory synapses; and an inhibitory component, $I\_in^r(t)$, produced by inhibitory synapses from the other interneuron.

The excitatory components, $I\_ex^r(t)$, targeting the auditory and visual interneurons are defined as:

$$\begin{aligned} I\_ex^{ia}(t) &= WI_a \cdot y^a(t) \\ I\_ex^{iv}(t) &= WI_v \cdot y^v(t) \end{aligned} \tag{6}$$

Where WI_a, WI_v are the weights of the excitatory connections from a unisensory input region to its corresponding interneuron element, assumed the same for both sensory modalities ($WI_a = WI_v = WI$).

The inhibitory input, $I\_in^r(t)$, that an interneuron receives from the interneuron of the other modality, through inhibitory synapses, is defined as:

$$\begin{aligned} I\_in^{ia}(t) &= L_{av} \cdot y^{iv}(t) \\ I\_in^{iv}(t) &= L_{va} \cdot y^{ia}(t) \end{aligned} \tag{7}$$

Where $y^{ia}(t)$ and $y^{iv}(t)$ are the activities of presynaptic auditory and visual interneurons respectively, and L_{av}, L_{va} are the strengths of the reciprocal inhibitory connections. These symmetrical synapses ($L_{av} = L_{va} = LI$) implement the WTA mechanism between the two areas. Also in this case, as in Eq. (5), we do not include a pure delay for the same reason stated above.

Multisensory output area — this region ($r = m$) receives a net input that is the sum of the stimuli, carried by long-range excitatory synapses, from the auditory and visual input areas.

Its net input, $ex^m(t)$, is defined as:

$$ex^m(t) = \sum_r Wm_r \cdot y^r\left(t - \Delta t^m\right), r = a, v; \tag{8}$$

Where $Wm_a = Wm_v = Wm$ are the weights of the excitatory connections from the unisensory input regions to the multisensory area and Δt^m is a delay, which represents the slightest latency with which stimuli from the input regions are able to generate behavioral responses.

For the sake of simplicity and to reduce the number of model assumptions, all the synapses previously described are symmetrical for the two sensory modalities.

Dynamics of each input component — All previous quantities (Eqs. (3) to (8)) affect the input $u^r(t)$ of the corresponding postsynaptic element via a second order differential equation. By denoting with $o_i(t)$ the output of the differential equation for the generic input source $i(t)$ (described by any of Eqs. (3) to (8)) we have

$$\begin{cases} \frac{d}{dt} o_i(t) = \delta_i(t); \\ \frac{d}{dt} \delta_i(t) = \frac{G_i^r}{(\tau_i^r)^2} i(t) - \frac{2 \cdot \delta_i(t)}{\tau_i^r} - \frac{o_i(t)}{(\tau_i^r)^2}; \end{cases} \tag{9}$$

Where G_i^r represents the gain and τ_i^r defines the time constant of the dynamics, for each region, r, and input component, i (Eqs. (3)–(8)). Eq. (9) implements a second-order impulse response with two coincident real poles. This is used frequently in neural modeling to mimic synaptic dynamics (see Cuppini et al., 2014; Jansen & Rit, 1995; Wendling et al., 2002). In the model, in order to reduce the number of parameters, we choose the same values for G_i^r and τ_i^r, for every connection (see Table 1), except two cases: (1) the external stimuli, and, (2) the feedback synapses implementing the cross-sensory inhibitory mechanism.

According to the previous description, the total input (say $u^r(t)$) received by a neuron in region r, is computed as follows: (1) for the input regions, is the sum of the external component Eq. (3), cross-modal term Eq. (4) and inhibitory feedback Eq. (5), filtered through the second order equation (Eq. (9)),

$$u^r(t) = o_e(t) + o_c(t) + o_l(t); \; with \; r = a, v \tag{10}$$

(2) for the inhibitory interneurons, it is the sum of the excitation from the input region (Eq. (6)) and the effect of the WTA mechanism (Eq. (7)), filtered by Eq. (9),

$$u^r(t) = o_{l\_ex}(t) + o_{l\_in}(t); \; with \; r = ia, iv \tag{11}$$

(3) for the output region, is the effect of the feedforward excitatory synapses (Eq. (8)), filtered by the differential equation previously described (Eq. (9)),

$$u^m(t) = o_{ex}(t). \tag{12}$$

2.3. Parameter assignment

The value of all model parameters (see Table 1) has been assigned from data present in the literature according to the main criteria summarized below.

Parameters of individual neurons — The central abscissa, θ, was assigned to have negligible neuron activity in basal conditions (i.e., when the input was zero). The slope of the sigmoidal relationship, s, was assigned to have a smooth transition from silence to saturation in response to external stimuli. The time constant agreed with values (a few ms) normally used in deterministic mean-field equations (Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Treves, 1993).

External input $e^r(t)$ — Physiological evidence shows that in the brain, auditory processing is faster, and auditory cortical neurons exhibit shorter latencies (e.g. Recanzone, Guard, & Phan, 2000) than neurons in the visual cortex (Maunsell & Gibson, 1992). As we did in a previous model (Cuppini et al., 2014), in this network the visual input region receives external stimuli described by a slower time constant, compared with the auditory ones. This is mimicked by setting $\tau_e^a < \tau_e^v$ in Eq. (9). The values of parameters τ_e^a and τ_e^v have been assigned to reproduce the temporal evolution of the process of an auditory and a visual stimulus in the early cortical areas. In particular, τ_e^a is given so that the auditory processing presents the faster dynamics, and the auditory area is activated by an auditory input 25–30 ms after the stimulus. Since two time constants represent the time needed for the activity in the input regions to reach 90% of its steady-state level, in response to a step input, we assume $\tau_e^a =$ 15 ms. For what concerns the visual area, τ_e^v is assigned so that a visual stimulus produces a detectable response in the visual area 45–50 ms after its onset; hence we choose $\tau_e^v = 25$ ms. It is worth noting that these are the only differences between the two sensory processing pathways (auditory and visual); all other parameters are assumed equal for the auditory and visual branches of the network, in order to reduce the number of ad hoc assumptions.

The strength of the external visual and auditory stimuli (parameters I_0^v and I_0^a) are chosen so that the overall input elicits a response, in the input regions, in the upper portion of the linear part of the sigmoidal static characteristic (i.e., a little below saturation).

Inhibitory component $l^r(t)$ — Behavioral data (Crosse, Foxe, & Molholm, 2019) show that strongest cross-sensory inhibitory effect occurs for ISIs as short as 1000 ms, but this effect decays slowly for longer time intervals between the stimuli. To simulate this result, we assumed that inhibitory mechanisms in the model have slow dynamics, implemented by time constants for the feedback projections as great as to 180 ms ($\tau_l^a = \tau_l^v$). In fact, with such a time constant, the inhibitory component provides a significant contribution to the input regions of the other modality after almost 1000 ms after stimulus presentation. In summary, the input regions are activated by an external stimulus after almost 50 ms (time constants 15–25 ms); through the excitatory projections, the interneurons show a non-null activity after 100–120 ms, and a peak activation between 250 ms and 300 ms; then the chosen inhibitory dynamics add 540 ms (approximately three time constants) before the feedback inhibitory component reaches its maximal effect on the unisensory input regions.

Cross-modal component $c^r(t)$ — The parameters of cross-modal components are selected to reproduce empirical findings by Raij et al. (2010). These authors, combining MEG and fMRI recordings, studied cross-modal activations and audio–visual interactions in the primary auditory cortex (i.e., A1) and in V1 at very early poststimulus latencies. To simulate Raij et al.'s (2010) results we used the input to the neurons (i.e., quantity $u(t)$ in Eq. (1)) since this is

indicative of field potentials, detected through EEG or MEG techniques, and/or synaptic metabolic activity, detected through fMRI. The dynamics of the cross-modal term are chosen symmetrical for both sensory modalities. Their time constant ($\tau = 15$ ms) and the delay in cross-modal synapses, $\Delta t = 16$ ms, simulating the latency with which the influence of a unisensory stimulus was detected in the area processing the other sensory modality, are selected so that the cross-modal component produces an effect on "the other region" after further 30–40 ms.

The efficacy of these synapses, *W*, has been set so that an activity elicited in one region by an external unisensory stimulus is not able by itself to activate the other sensory region. However, it is strong enough so that, in case of a multisensory stimulation, the effect of the excitatory cross-modal synapses enhances the level of activity in the opposite input area.

These parameters help to simulate a multisensory interaction, occurring between unisensory input regions, producing a rapid transient excitatory effect between these input areas, as mounting evidence in the literature suggests happens as early as the primary cortices (see Alais, Newell, & Mamassian, 2010; Driver & Noesselt, 2008; Foxe & Schroeder, 2005; Ghazanfar & Schroeder, 2006; Musacchia & Schroeder, 2009; Recanzone, 2009; Shams & Kim, 2010; Stein & Stanford, 2008; Ursino, Cuppini, & Magosso, 2014 for reviews). This effect has a rapid time course as described above. This mechanism produces a greater activation of the input layer in the case of multisensory presentations.

For the next three elements the time constants are the same and fixed as the faster dynamics previously chosen for the auditory and cross-modal synapses, $\tau_i^r = 15$ ms; *with* $i = I_{ex}, I_{in}, m$.

Interneurons excitatory components, $I\_ex^r(t)$ — what characterizes this input is the efficacy of the excitatory synapses from the unisensory input regions, *WI*. Its value is chosen so that even a small activity in the input layer is able to activate the cross-sensory inhibitory mechanism, by eliciting an activity in the corresponding interneuron.

Interneurons reciprocal inhibitory input, $I\_in^r(t)$ — this element implements the WTA competition. The effectiveness of the reciprocal synapses, *LI*, is chosen high enough so that the "winner" interneuron is able to turn off almost completely the "loser" element.

Excitatory input to the multisensory area, $ex^m(t)$ — the behavior of the model is evaluated by comparing the simulated RTs to the behavioral data reported in the literature. To simulate RTs, we compare the elicited activity in the multisensory area with a "detection threshold", chosen equal to 30% of the maximum value of such activity, as described in the manuscript's method section. The strength of the excitatory feedforward synapses, *Wm,* is chosen so that even a small activity (i.e., 30% of its saturation value) in the input regions, visual and auditory, elicits a response in the output area in the upper portion of the linear part of the sigmoidal static characteristic (i.e., a little below saturation). The delay Δt^m (= 100 ms) has been assigned in accordance with the threshold assumed by Crosse, Foxe, and Molholm (2019) to discriminate fast outliers (quicker RTs, < 100 ms, are considered anticipatory responses). These synaptic elements are intended to simulate the neural process required for an external input to produce a motor response in the output region.

For the above elements, the G_i^r values are chosen so that the elicited activity in the post-synaptic elements is in the linear portion of the sigmoidal relationship: as shown in Table 1, $G_i^r = 75$ for every synapse ($i = e, c, I\_ex, I\_in, m$), except for the inhibitory feedback ($i = l$) where $G_l^r = 750$.

2.4. Assessment of network performance

First, as described above, to discriminate clearly between alternative architectures for the network, we performed an analysis of the data collected by Crosse, Foxe, and Molholm (2019), based on the duration of the ISIs, in the case of Repeat and Switch conditions, and for unisensory and multisensory stimuli. Once we identified inhibition as the more plausible neural architecture to explain switch effects in Crosse's data, we next performed several simulations to test the network behavior and identify the most important neural and architectural mechanisms implemented in the model (e.g., feedforward connections, cross-modal synapses, inhibition from competitive layer). To this end, we presented sequences of unisensory (auditory-alone and visual-alone) and multisensory stimuli (audiovisual inputs) in randomized order at pseudo-randomly chosen ISIs between 1000 ms to 3000 ms (using a boxcar function). Simulated RTs are computed as the time interval between the instant of input presentation and the instant when the evoked activity in the output area reaches threshold. These results were analyzed separately based on the respective input modality. Furthermore, we discriminated between "Repeat" trials (the preceding stimulus belonged to the same sensory modality) and "Switch" trials (the preceding stimulus was of a different sensory modality). As described above, in the network, the inputs used in each repetition, for each stimulus condition, were randomly chosen from a uniform distribution in order to replicate the within-subject variability of sensory stimuli in a real environment. Therefore, for every stimulus configuration, first we computed the mean RTs, obtained from 100 simulations with the same input condition, then, we compared these model's RTs with the mean RTs extracted from the adult population from Crosse et al.

Finally, to simulate inter-subject variability, i.e. the behavioral differences among the subjects involved in the Crosse experiment, in the last set of simulations, we introduced a further noise component to the parameters describing the different architectural mechanisms implemented in the model. This allowed the simulation of different subjects, characterized by different values of specific parameters. As described before, for each simulated subject, we evaluated mean RTs for every stimulus configuration (unisensory/multisensory and repeat/switch conditions), over 100 presentations.

3. Results

In the following, we present results describing the network's behavior under different scenarios. These allow us to understand (1) the effects of unisensory switch conditions, compared with the repeat configuration; (2) the differences in the case of multisensory presentation; and (3) the influence of ISI (short versus long) on the network's performances. Finally, (4) we discuss which of the network elements are responsible for the inter-subject variability observed in the empirical data.

3.1. Switch vs repeat for unisensory trials

First, examples of simulations of unisensory trials are presented.

Figs. 3 and 4 compare the evoked activities in the regions of the model in case of Auditory Repeat and Auditory Switch conditions.

The auditory stimuli have the same efficacy, and in both conditions the second stimulus (Auditory) is presented with an ISI = 2000 ms. As it is evident in Fig. 4, in the case of the Switch condition, the activity in the auditory region (green lines in figures) evoked by the second stimulus is lower than activity in

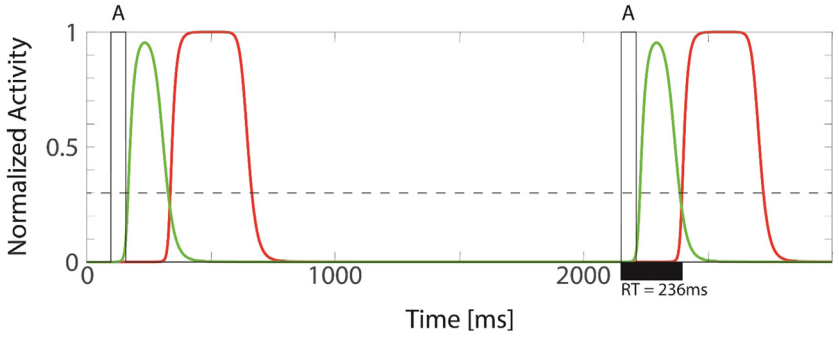

Fig. 3. Example of the network behavior, in the case of an Auditory Repeat Condition. Black line represents the external auditory stimuli presented with an ISI of about 2000 ms. Green line describes the activation of the auditory region. Red line represents the activity elicited in the output region in response to this stimulation. This activity is compared with the detection threshold, the dashed line, to compute the simulated RT, horizontal black bar in figure. In this case RT = 236 ms. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

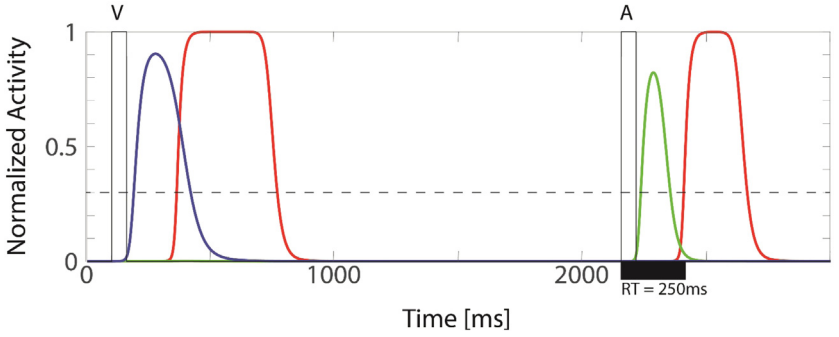

Fig. 4. Example of the network behavior, in the case of an Auditory Switch Condition. Black line represents the external visual and auditory stimuli presented with an ISI of about 2000 ms. Blue and green lines describe the activation of the visual and auditory regions respectively. Red line represents the activity elicited in the output region in response to this stimulation. This activity is compared with the detection threshold, i.e., the dashed line: the simulated RT, horizontal black bar, is calculated as the interval between the instant when the red line overcomes the threshold and the instant of stimulus presentation. In this case RT = 250 ms. As expected, in Switch condition, the RT is longer than the RT in Repeat condition.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the Repeat case. This reduced activation is produced by the inhibitory interneuron excited by the preceding visual stimulus. In the model, the efficacy of this cross-modal inhibition over a long period is simulated by means of the slow dynamics used to implement the effect of the inhibitory synapses. Then, the lower activation of the auditory region produces a delayed activation of the multisensory/motor region, with a consequently longer RT.

Similar results were found in case of the visual switch condition: a visual input, preceded by an auditory stimulus, would elicit a longer RT compared to the visual repeat condition.

As shown in Figs. 5 and 6, the preceding auditory input lowered the overall activity elicited in the visual region by the external stimulation, thus resulting in a slower activation of the multisensory area and in a longer RT in response to the visual input. As described for the auditory switch/repeat conditions, this effect and the related visual switch cost is produced by the auditory driven long-lasting inhibition to the input visual region.

3.2. Influence of ISI

As suggested by the previous results, the unisensory Switch/Repeat conditions highlight the potential role played by cross-modal inhibition when the brain operates in a multisensory environment, where stimuli of different sensory modalities, in different temporal arrangements, must be processed. As stated in the Introduction, an important question is the effect of ISI on the RTs in the case of the Switch condition. The expectation is that short ISIs would produce a greater Switch cost compared to long ISIs, assuming inhibition as the predominant mechanism. This expectation was confirmed by the statistical analysis performed on the empirical data, as discussed above. Here, we further test this hypothesis by stimulating the model with unisensory auditory and visual inputs, using with different ISIs.

Fig. 7 describes the effect of different ISIs in case of Auditory Switch. In case of fast ISIs (panel A), it is evident that the inhibition of the preceding visual input is stronger on the auditory input, compared to the condition reported in Fig. 4, eliciting a lower activity in the auditory area and a slower response in the multisensory region. The opposite behavior has been found in the case of a long ISI between the two stimuli: as shown in panel B, Fig. 7, in this case the activity elicited by the second auditory stimulus is stronger, compared to the condition depicted in Fig. 4, revealing a lower inhibition produced by the visual interneuron.

Based on these results we can infer that the longer RTs for the switch versus repeat conditions, and the correlated switch cost, is due to the cross-modal inhibition produced by the preceding stimulus, and that the inhibition effect decreases with increasing ISI.

3.3. Unisensory vs multisensory stimulation

The previous results help describing the behavior of the brain in the case of unisensory Switch/Repeat conditions. However, empirical results suggested that different mechanisms are recruited in the case of multisensory stimulation. The analysis performed by Crosse et al. suggests that multisensory stimuli are not affected by the preceding conditions: multisensory repeat versus switch

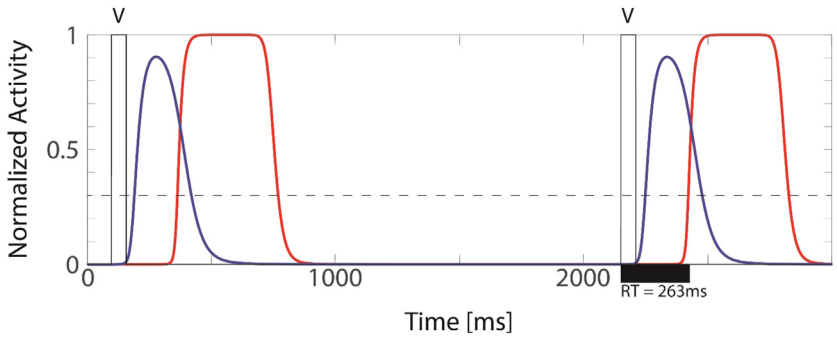

Fig. 5. Example of the network behavior, in the case of a Visual Repeat Condition. Black line represents the external visual stimuli presented with an ISI of about 2000 ms. Blue line describes the activation of the visual region. Red line represents the activity elicited in the output region in response to this stimulation. This activity is compared with the detection threshold, the dashed line, to compute the simulated RT, horizontal black bar. In this case RT = 263 ms. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

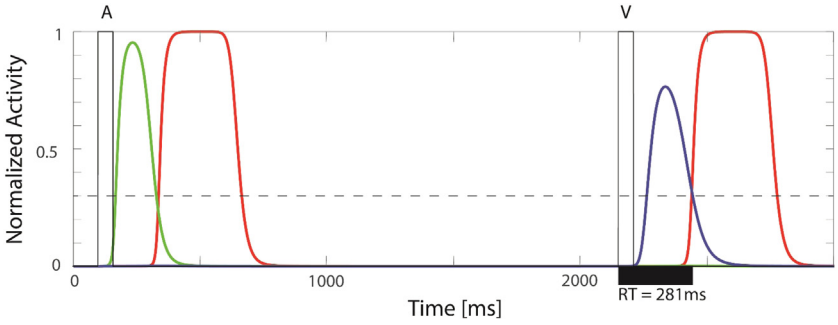

Fig. 6. Example of the network behavior, in the case of a Visual Switch Condition. Black line represents the external auditory and visual stimuli presented with an ISI of about 2000 ms. Blue and green lines describe the activation of the visual and auditory regions respectively. Red line represents the activity elicited in the output region in response to this stimulation. This activity is compared with the detection threshold, the dashed line, to compute the simulated RT, horizontal black bar. In this case RT = 281 ms. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

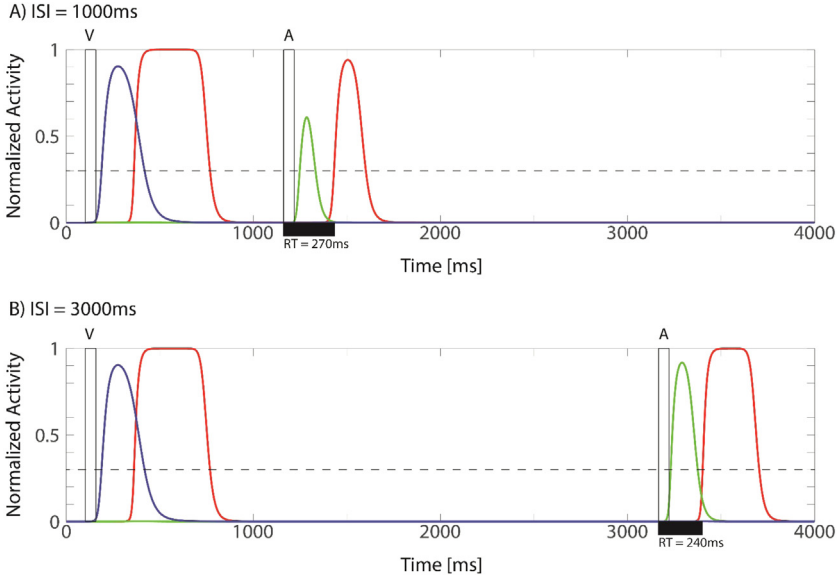

Fig. 7. Example of the ISI's effect on the network behavior, in the case of an Auditory Switch Condition. Black line represents the external auditory and visual stimuli. Blue and green lines describe respectively the activation of the visual and auditory regions. Red line represents the activity elicited in the output region in response to this stimulation. This activity is compared with the detection threshold, the dashed line, to compute the simulated RT, horizontal black bar. (A) With an ISI of 1000 ms, the model responds with a RT = 270 ms. (B) With an ISI of 3000 ms, the effect of the preceding visual input is lower and the model presents a RT = 240 ms. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

RTs did not differ significantly ([mean ± SEM] RTs in Crosse, Foxe, and Molholm (2019): AV repeat = [260.7 ± 7.6] ms, AV switch = [264.8 ± 7.6] ms). Stimulating our model with these stimulus configurations, we were able to demonstrate that, as expected, the likely underlying neural mechanisms recruited in the case of multisensory stimuli and those responsible for the behavioral

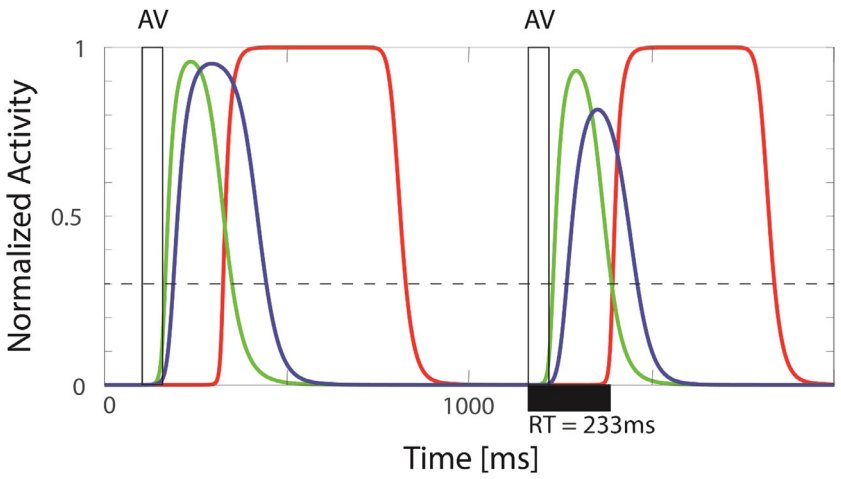

Fig. 8. Audiovisual (AV) Repeat Condition. Black line represents the external AV stimuli. Blue line describes the activation of the visual region, the green line represents the auditory activity. Red line represents the activity elicited in the output region in response to this stimulation, and it is compared with the detection threshold, the dashed line. The stimuli were presented with an ISI of about 1000 ms, which is the case with strong inhibitory influence of the preceding stimulus on the second one. As we can see in the figure, the activities elicited in the input regions by the second AV stimulus are slightly depressed, compared with the first activation, as result of the cross-modal inhibition. Nevertheless, the multisensory region is highly stimulated and its activity reaches the saturation level. The computed RT, horizontal black bar, in this case, is RT = 233 ms. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

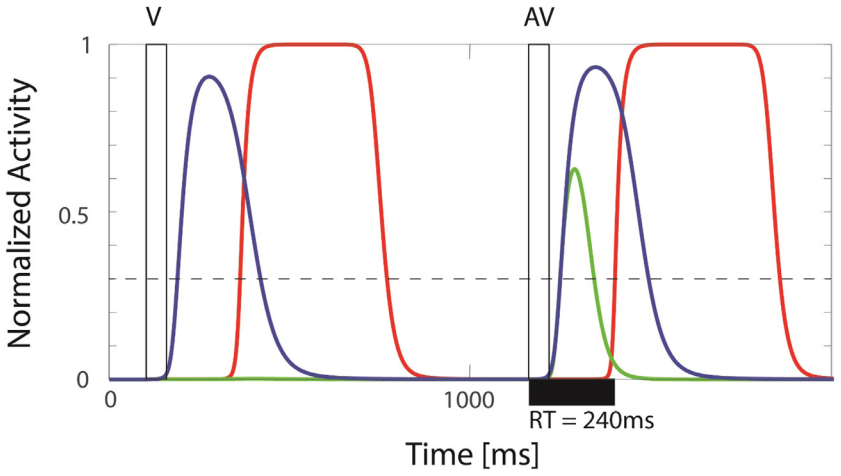

Fig. 9. Audiovisual (AV) Switch Condition. The AV input is preceded by a visual stimulus, that exerts an inhibitory effect on the auditory component of the following AV stimulus. Nonetheless, also in this configuration the multisensory area receives a strong excitation, resulting in an activity comparable with the AV repeat condition, and similar RTs (in this case, is RT = 240 ms, horizontal black bar). . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results in the case of switch and repeat conditions are the cross-modal connections between the unisensory input regions and the feedforward projections converging to the output area.

Figs. 8 and 9 show the evoked activities modeled in each sensory region in the case of multisensory repeat and switch conditions, respectively.

From the comparison between these two conditions, we can see that the preceding stimulus has an inhibitory effect on the following AV input in both conditions: the peak activities in the input regions (green and blue lines) are lower for the second presentation. However, this effect is compensated for by the excitatory cross-modal connections between the two unisensory input regions and by the convergent afferents to the multisensory area, so that in both conditions, AV repeat and AV switch, the effect of the preceding stimulus on the RTs is nullified by the cross-modal arrangement of the network. These two mechanisms implement the multisensory facilitation/enhancement that produces a stronger activity in the output region, and therefore quicker RTs, in response to AV stimuli, compared to the unisensory stimulations. This architectural implementation matches with our previous studies and computational models realized to simulate MSI in various experimental conditions and for different perceptual and cognitive processes. For examples, the role played by cross-modal direct connections between sensory primary regions has been found to be critical in case of multisensory illusions, as spatial ventriloquism, sound-induced flash illusion and fusion effect (Cuppini et al., 2014, 2017a; Magosso et al., 2012; Ursino, Cuppini, & Magosso, 2017; Ursino et al., 2019); the feedforward synapses, converging to a multisensory region, are critical to explain the integrative abilities of the Superior Colliculus and their maturation (Cuppini et al., 2018; Cuppini et al., 2011; Cuppini et al., 2012); and both mechanisms, and their maturation through specific multisensory experience, simulate and explain how the brain deals with processes of a higher level of complexity, such as the solution of the Causal Inference (Cuppini et al., 2017a) and the acquisition of MSI language abilities (Cuppini et al., 2017b).

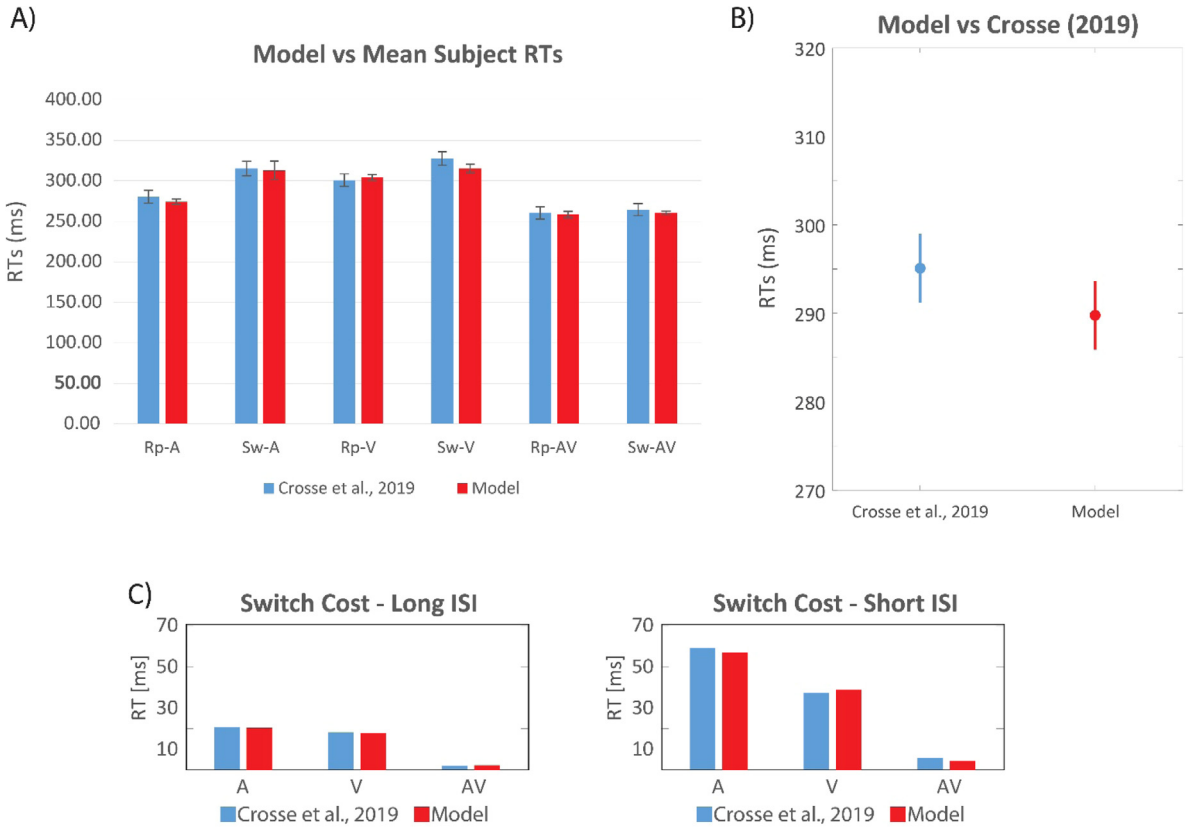

Fig. 10. Simulated RTs vs Mean RTs of the subjects' population. In figure (A), for each stimulus configuration (Auditory and Visual Repeat, Auditory and Visual Switch, AV Repeat and Switch) we compared RTs obtained with the model basal configuration with mean RTs (plotted with their SEM) extracted from the population of subjects. (B) RTs collapsed across all conditions obtained by model simulations and from Crosse et al. 2019. Results show that the simulated RTs and the Mean Subjects' behaviors are statistically comparable. (C) Effect of the ISIs on the different stimuli configurations: we evaluated the simulated Switch Costs in the unisensory and multisensory cases, for short and long ISIs, and compared with those computed from Crosse's data. The model shows behaviors and results comparable with the empirical data.

3.4. Mean behavior of the model

Once we identified the main mechanisms involved, the next step in our analysis is to determine the exact parameters' values of such mechanisms that allow the simulation of the mean RTs computed from the empirical data of Crosse et al..

As shown in Fig. 10A, the model in its basal configuration provides a good simulation of the empirical data from adults in Crosse, Foxe, and Molholm (2019). Supporting a good fit between the model and the empirical data, a 3-way ANOVA with factors of data source (model vs empirical), condition (repeat vs switch), and sensory modality (A, V, AV) was performed. This analysis showed significant main effects of condition ($F_{1/483} = 21.95$, $p < 0.0001$) and sensory modality ($F_{2/483} = 63.82$, $p < 0.0001$), but not for source (Fig. 10B, $F_{1/483} = 1.82$, $p = .178$). What is more, source did not interact with either modality or condition. These results suggest a good fit between the model simulations and the experimental data.

Finally, as it was our initial hypothesis, the model's results display a strong dependence on ISIs: as shown in the empirical data, also the simulations exhibited a more effective inhibition, for the Switch condition, in case of short ISIs. As described in Fig. 1C, this effect, evaluated in terms of Switch Cost, is significant only in the case of unisensory stimuli, and not in the case of AV configurations. Moreover, the Switch Costs computed for the model's results, fit well with the empirical data, from Crosse et al., as shown in Fig. 10C.

3.5. Inter-subject variability

Crosse et al., in their analysis, found significant switch costs only for the unisensory conditions, auditory and visual stimuli; but they revealed also that individual subjects, belonging to the same age-group, presented very different RTs in response to the same input configuration and wide inter-subject variability for the inhibitory effect of the cross-sensory competition, as indexed by the switch cost.

From the basal model configuration, to identify which neural mechanisms were responsible for the variability observed amongst the analyzed subjects, we modified randomly the value of single parameters and compared the results of simulations with the experimental data of the overall population, in case of auditory and visual repeat and switch conditions, and in the case of multisensory stimuli.

Among the mechanisms implemented in this model, only two had a strong effect in simulating the RTs in the different conditions: (1) the feedforward connections, which carry the information extracted from the input regions to the motor/multisensory output area, and (2) the cross-modal inhibition, which is mediated by the inhibitory interneurons.

In contrast, our modeling suggests that cross-modal synapses, the WTA mechanisms, and the time constants do not affect the inter-subject variability of unisensory RTs.

In the following section, we discuss the results obtained by modifying the parameters characterizing these mechanisms.

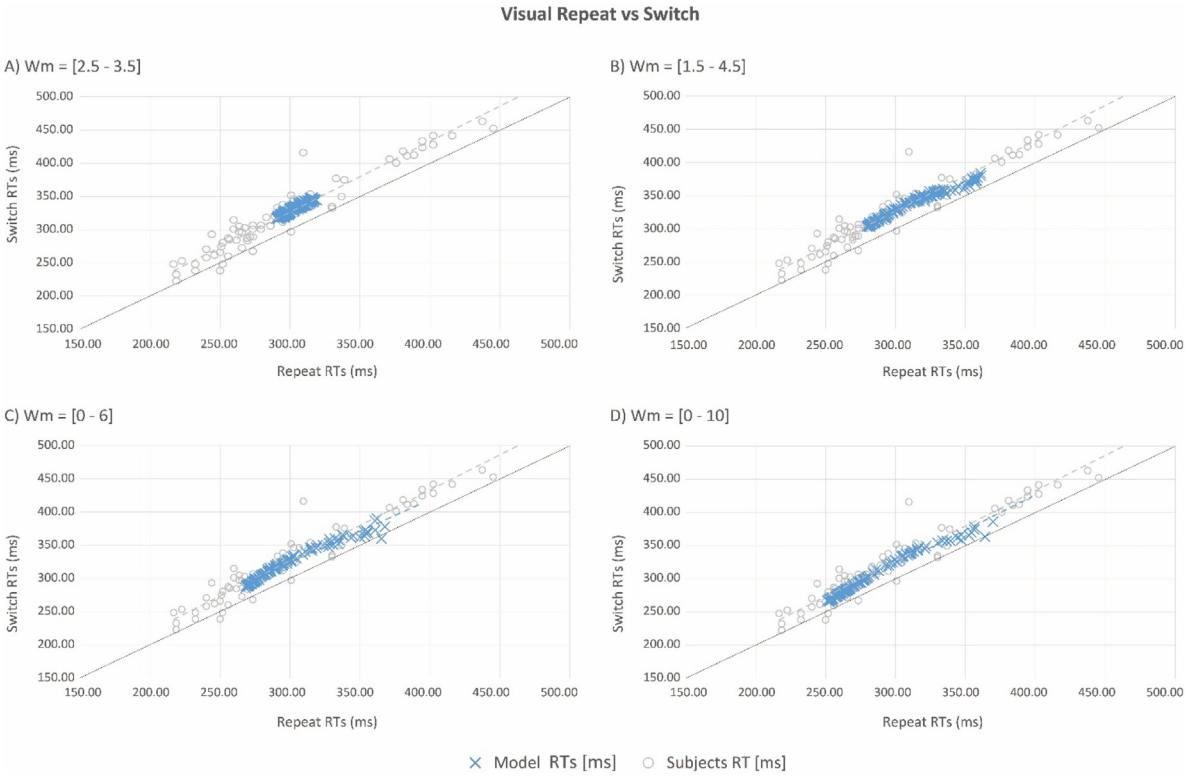

Fig. 11. *Wm* (feedforward effectiveness) contribution to inter-subject variability. Simulated mean RTs are compared with each subject's mean RTs, in the case of Visual Repeat versus Switch configurations. The 4 panels differ only for the range of effectiveness of the feedforward connections. The *x*-axis reports the RTs in the case of the Repeat condition, the *y*-axis the RTs to the Switch condition. Blue 'Xs' represent the mean RTs of the "simulated" subjects, gray circles the subjects' mean RTs, from Crosse, Foxe, and Molholm (2019). The diagonal black line in each panel represents conditions in which switch RTs are equal to repeat RTs (no switch cost); thus, the vertical distance between data points and the diagonal represents the mean switch cost for each "simulated" or "real" subject.

Feedforward excitatory connections

These connections are characterized by two parameters: effectiveness and time constants. Only by modifying the first were we able to produce inter-subject RT variability comparable to that observed in the adult subjects. A change in the effectiveness of these synapses was realized by adding a random component to the basal value: this component was generated from a uniform distribution with zero mean and maximum and minimum bounds equal to $\pm$ an assigned percentage of the basal value ($Wm = Wm0 \pm$ noise%). Each value of this parameter can be considered representative of a different simulated subject. In order to characterize the individual behavior, the mean RTs were calculated in each simulated subject, and for all stimulus configurations, over 100 repetitions per condition. To simulate a physiological input variability among these repetitions, the strengths of the auditory and visual input stimuli in each individual subject were chosen from a uniform distribution (as described previously).

An example of the effect of varying the feedforward effectiveness is reported in Fig. 11, where we compared the experimental data from Crosse et al. to the model's results, in the case of visual repeat and switch conditions: each panel was obtained by varying the range of values for the parameter. All the other parameters were set at their basal values. Similar results can be obtained in the case of auditory repeat and switch conditions, but are not reported here for brevity.

However, as can be observed in Fig. 11, a change in the strength of feedforward connections can only partially reproduce the inter-subject RT variability: hence, we added a second random component to model parameters: the effectiveness of the feedback inhibition, as described below.

Inhibitory feedback

The second parameter analyzed was the inhibition that neurons exert on the input regions through their inhibitory connections. As described in the methods section, this inhibition is mediated by two synapses: the excitatory connections from the input regions to the "*competitive layer*"*(WIa and WIv)*, and the feedback inhibitory synapses from this layer to unisensory input regions (L_a and L_v, see Fig. 2). Both connections are characterized by their effectiveness and time constant. As in the previous case, simulations show that the time constants have little effect on the inter-subject variability, whereas the effectiveness of both excitatory and inhibitory synapses in this inhibitory network is meaningful. More precisely, the product of the two synaptic weights, i.e., the overall feedback strength, accounts for the inhibitory effect.

As done for the feedforward connections, to test the role played by inhibitory synapses we added a random component (noise%) chosen from a uniform distribution to their basal value. Then we simulated a network where both random components (effectiveness of feedforward connections and effectiveness of inhibitory connection) can vary randomly. As an example, Fig. 12 shows the effect of the manipulation of the feedback inhibitory synaptic weight on the switch cost in the case of visual stimulation, compared with the empirical results from Crosse et al.. These results were obtained by increasing the range of values of the inhibitory connection, while the random value of the effectiveness of the feedforward connection was chosen in the range [0–10] (the range–value used in Fig. 11D). All other parameters were set at their basal value.

As expected, a fluctuation in the efficacy of inhibition strongly affects the RTs in case of unisensory switch conditions, and helps explain the differences found among subjects. As it is clear from

Fig. 12. Inhibition contribution to inter-subject variability. Simulated mean RTs are compared with each subject's mean RTs, in the case of Visual Repeat versus Switch configurations. The 4 panels differ only for the range of effectiveness of the inhibitory feedback connections. In each panel, the synapses *L* of the inhibitory feedback were randomly varied (uniformly sampled) within a different range, while the feedforward connections *Wm* were varied within the same range [0 -10]. The *x*-axis reports the RTs in the case of the Repeat condition, the *y*-axis the RTs to the Switch condition. Blue 'Xs' represent the mean RTs of the "simulated" subjects, gray circles the subjects' mean RTs, from Crosse, Foxe, and Molholm (2019). The diagonal black line in each panel represents conditions in which switch RTs are equal to repeat RTs (no switch cost); thus, the vertical distance between data points and the diagonal represents the mean switch cost for each "simulated" or "real" subject.

Fig. 12, a random component in the inhibitory loop dramatically increases the inter-subject variability especially in terms of the difference between switch and repeat RTs, i.e., the so-called switch cost.

Additional simulations, not presented for brevity, show that variability in other parameters of the model have only minor effects on the simulated behaviors in any of the analyzed configurations of the stimuli.

4. Discussion

The model described in this work aspires to reproduce the experimental results by Crosse, Foxe, and Molholm (2019) with minimal complexity and a high-level of abstraction. For this reason, just three mechanisms have been included:

(i) The unisensory areas A and V exchange reciprocal cross-modal excitatory connections. These connections implement a facilitation mechanism, which is the basis of many important results concerning multisensory integration. In particular, in the present paper we simulated only audio–visual stimuli at a single specific spatial position (and so we used just a single neural unit per area, representing a population of neurons that codes for that position). Previous modeling work, including more neurons to code for different azimuthal positions (Cuppini et al., 2017a; Ursino et al., 2019) demonstrated that cross-modal excitatory connections are at the basis of some illusory phenomena, such as the spatial ventriloquism (Cuppini et al., 2017a; Magosso, Cuppini, & Bertini, 2017; Magosso et al., 2012) and the sound-induced flash illusion (Cuppini et al., 2014). Moreover, previous studies demonstrated that these synapses can be trained by experience, to implement the prior probability of the co-occurrence of audio–visual stimuli in close temporal and spatial proximity (Ursino et al., 2017). Various data in the literature confirm the existence of cross-modal links between the primary visual and auditory regions ((Alais et al., 2010; Driver & Noesselt, 2008; Foxe & Schroeder, 2005; Ghazanfar & Schroeder, 2006; Musacchia & Schroeder, 2009; Recanzone, 2009; Shams & Kim, 2010; Stein & Stanford, 2008; Ursino et al., 2014) for reviews). Finally, we wish to remark that these cross-modal synapses in the model are not strong enough to evoke a phantom response in the other area in the case of unisensory inputs, i.e., only a single area is active in the unisensory condition.

(ii) A second facilitatory mechanism in the model is implemented by the feedforward synapses converging from the auditory and visual areas towards a multisensory area, where the constituent unisensory signals are fully integrated and the behavioral response is elicited. In particular, the large temporal delay used for these connections wishes to represent, although schematically, not only the delay of neural transmission, but also the time required to produce a motor activity. It is worth noting that the presence of a sigmoidal relationship in the multisensory region allows several well-known characteristics of multisensory integration to be reproduced, such as the enhancement and the inverse effectiveness (see Cuppini et al., 2018, 2010; Cuppini et al., 2011; Cuppini et al., 2012; Magosso et al., 2008; Ursino et al., 2009a).

In summary, mechanisms 1 and 2 realize the classic schema of multisensory integration. In this regard, it is worth-noting that both mechanisms have been implemented with a small synaptic time constant (15 ms), of the same order as that used to characterize the auditory response: hence, they work to integrate

auditory and visual stimuli in close temporal proximity. Indeed, it is well known that classic multisensory facilitation occurs only if stimuli of different sensory modalities happen within a short temporal window (*TWI* = 40 to 400 ms, depending on context. See Meredith et al. (1987), Stein and Meredith (1993) and Miller et al. (2015) for analysis at neural level, and behavioral results from Bell et al. (2005, 2006), Colonius and Diederich (2004), Lewald et al. (2001), Lewald and Guski (2003), Mégevand et al. (2013), Meredith (2002), Musacchia and Schroeder (2009), Navarra et al. (2005), Romei et al. (2007), Rowland and Stein (2007, 2007), Spence and Squire (2003), Stevenson and Wallace (2013), van Wassenhove et al. (2007) and Wallace et al. (2004).

(iii) In order to simulate the switch cost observed by Crosse, Foxe, and Molholm (2019), when a stimulus temporally precedes a second stimulus of a different sensory modality, we added a third mechanism to the previous classic schema: a competition (mediated by inhibitory synapses) between the two modalities. Indeed, as considered in the Introduction, the switch cost can be understood assuming a long-lasting inhibition of the first stimulus over the second. This is characterized by a much longer time constant (180 ms) than facilitatory multisensory effects and, with the protracted dynamics involved in a feedback competitive loop, it operates with a time scale compatible with the RTs observed by Crosse, Foxe, and Molholm (2019).

It is worth noting that the inhibitory mechanism plays a relevant role only in the presence of sequential unisensory stimuli of different sensory modalities (A–V or V–A), and that this mechanism is in action for stimuli separated by a 1 s interval. The lower limit of the inhibitory effect is not yet clear. In Crosse, Foxe, and Molholm (2019), this limit has not been tested, and we are not aware of similar studies in the literature directly investigating competition among sensory modalities in the temporal domain. Our model is an attempt to formulate a theoretical framework, based on the available data on multisensory processing, temporal integration window and inhibition among sensory modalities, that can be used to formulate testable predictions to shed light on this cross-sensory interaction. For example, with this structure, and the chosen parameter values, the model predicts that this inhibitory effect reaches its peak at SOAs of about 1s. Further experiments, in which a larger range of SOAs is tested (including very short SOAs, where MSI is expected to dominate) will be needed to fully characterize this inhibition function. Additional model simulations, not showed here, further show that, for ISIs < 400 ms, the time course of cross-modal facilitation, mediated by the dynamics of the feedforward and cross-modal synapses, overtakes the effect of the inhibition, resulting in the classic multisensory integration reported in the literature. This 400 ms window fits with the upper bounds of the TWI for MSI (Bell et al., 2005, 2006; Colonius & Diederich, 2004; Lewald et al., 2001; Lewald & Guski, 2003; Mégevand et al., 2013; Meredith, 2002; Musacchia & Schroeder, 2009; Navarra et al., 2005; Romei et al., 2007; Rowland & Stein, 2007, 2007; Spence & Squire, 2003; Stevenson & Wallace, 2013; Wallace et al., 2004; van Wassenhove et al., 2007).

Since data by Crosse, Foxe, and Molholm (2019) exhibit large individual variability, we further tested the model by investigating the possibility to emulate inter-subject differences, acting on the two parameters which mainly affect the temporal responses: i.e., the strength of the feedforward synapses from the input layer to the motor output layer, and the overall strength of the feedback inhibitory loops. Our tests show that the first mechanism can explain the large differences in the RTs observed from one subject to the other in the repeat conditions (V–V or A–A). The reason is that a response is elicited in the model only when activity in the multisensory area overcomes a given threshold (0.3 in the present simulations). Of course, the stronger the feedforward synapses, the shorter the time required for the multisensory activity to overcome the threshold.

Conversely, as expected, the strength of the feedback competitive mechanism mainly affects the difference between the switch and the repeat RTs, i.e., the so-called "switch cost", which also exhibits large differences among subjects.

While we did not explicitly model differences in inter-trial variability across conditions, this is inherently captured by the parallel neural architecture as a result of statistical facilitation (100 simulations per condition taken from a uniform distribution; Raab, 1962), as well as the enhanced temporal dynamics due to facilitative multisensory integrative processes. This resulted in the simulated AV distributions having less spread than the A and V distributions (see SEM, Fig. 10A), in line with previous observations (Innes & Otto, 2019; Otto, Dassy, & Mamassian, 2013).

An important aspect to be discussed concerns the function of the competitive mechanism in the present model, where it may be located in the brain, and the role of cross-modal integration mechanisms (facilitatory and inhibitory) in daily life. In everyday life, from the moment we wake to the moment we go to sleep, we encounter multiple sensory inputs that come at us in an everchanging stream. There are thus inherent switches in the sensory modality of the inputs we are attending to. Behaviorally, this has been shown to incur a "cost", even when all stimuli are task relevant and therefore presumably attended. Such switching is arguably the more natural state of an organism when interacting with the external environment, and yet the underlying neural mechanisms driving these so-called costs are not well understood. In particular, when dealing with multisensory information in daily life, two opposite conditions must be recognized: integration vs. separation. If two stimuli of different sensory modalities are in close spatial and temporal proximity, they must be integrated into a single percept, and ascribed to a single external cause. In this case, multisensory facilitation plays the major role. Conversely, stimuli occurring at greater temporal and/or spatial disparity should be treated as separated percepts, produced by different external causes. An inhibitory mechanism may thus be of behavioral value in this case, to focus attention onto the most relevant stimulus, while neglecting the other.

The model simulates the visual and auditory switch and repeat task. Hence, we can hypothesize that the competitive loop in our model is mainly part of an attention mechanism involved in cognitive control. Indeed, one classic way to study cognitive control is to ask participants to switch from one task to another and compare the relative performance. Various studies suggest that this switching has a cost: response speeds are typically slower and task accuracy poorer following a task-switch than following a task-repeat (Foxe, Murphy, & De Sanctis, 2014; Jersild, 1927; Koch & Allport, 2006; Rogers & Monsell, 1995; Spector & Biederman, 1976; Weaver et al., 2014; Wylie, Javitt, & Foxe, 2003b). The same applies to switching sensory modality during a multisensory RT paradigm (Gondan et al., 2004; Miller, 1982; Otto & Mamassian, 2012, 2017; Shaw et al., 2020; Spence, Nicholls, & Driver, 2001). More specifically, to investigate cross-modal interactions, some authors adopted a cross-modal cue–target paradigm, in which attending to a cue in one modality would delay the response in the other modality, especially if the temporal distance between the cue and the target is longer than a given stimulus onset asynchrony (SOA) or if a neural cue is interposed. This phenomenon has been called "cross-modal repetition inhibition" (Reuter-Lorenz, Jha, & Rosenquist, 1996; Wang, Yue, & Chen, 2012; Wu et al., 2019). We can thus speculate that the proposed neural circuit may be part of an attention control mechanism, which works to solve conflict and/or to improve flexibly by switching attention. This circuit is likely located in the frontal

and parietal lobes: in fact, prior studies have demonstrated that these regions are more active during a switch than during a task repetition (Braver, Reynolds, & Donaldson, 2003; Buchsbaum et al., 2005; Dove et al., 2000; Greenberg et al., 2010; Liston et al., 2006; Wylie, Javitt, & Foxe, 2003a, 2004; Wylie et al., 2009). It is worth noting that we used a similar inhibitory control loop to simulate conflict resolution in bilingualism (Cuppini, Magosso, & Ursino, 2013).

Other computational work that has successfully modeled multisensory RT distributions included an attentional component in the form of a channel dependency (free correlation parameter) which could potentially account for such a competitive mechanism (Innes & Otto, 2019; Otto et al., 2013; Otto & Mamassian, 2012). Future work is required to reconcile the cognitive and neural architectures of multisensory processing proposed in the computational literature (Crosse et al., 2019).

Finally, an interesting aspect of the experimental data, that the present model does not reproduce, concerns the difference between the short and long RTs in the unisensory repeat condition (about 20 ms, see Fig. 1). Indeed, this difference is not statistically significant, but may deserve a critical comment. One possible origin of this difference may be subject expectation (Niemi & Näätänen, 1981; Spence et al., 2001), which can increase for longer foreperiods, thus reducing the RT. Indeed, Niemi and Näätänen (1981) and Luce (1986) analyzing the effects of foreperiods on the RTs to sequences of stimuli, linked the speed-up RTs in case of long foreperiods to the increased temporal expectation of the stimuli. However, although this effect can contribute to the observed data, it cannot explain differences between the switch and repeat conditions. In fact, in case of switch conditions, the difference between the short and long foreperiods become significantly stronger, thus suggesting the presence of an additional mechanism than simple expectancy. We thus suggest that a pivotal role is played by a cross-modal inhibitory interaction, which becomes relevant only in case of a switch between two sensory modalities.

Further studies are necessary to validate the proposed model in the context of other conflict paradigms and to establish whether the present hypothetical mechanism can be used to explain attention, flexibility and switch cost in a larger variety of experimental data. Furthermore, the same circuit could be used to examine differences in attentional shift mechanisms in neurotypical versus clinical subjects, for instance in autism spectrum disorders (Crosse, Foxe, & Molholm, 2019; Williams, Goldstein, & Minshew, 2013). Crosse et al. already provide evidence to suggest the role of prolonged competitive interactions in ASD, and previously linked such empirical evidence to an inhibitory neural architecture such as the one proposed here (Crosse et al., 2019).

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.

Funding

This work was supported by the University of Bologna, Italy.

References

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3–38.

Bell, A. H., et al. (2005). Cross-modal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *Journal of Neurophysiology*, *93*, 3659–3673.

Bell, A. H., et al. (2006). Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. *Experimental Brain Research*, *174*(1), 53–59.

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(9), 3844–3848.

Bizley, J. K., & King, A. J. (2008). Visual–auditory spatial processing in auditory cortical neurons. *Brain Research*, *1242*, 24–36.

Bizley, J. K., & King, A. J. (2009). Visual influences on ferret auditory cortex. *Hearing Research*, *258*(1–2), 55–63.

Brandwein, A. B., et al. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex*, *21*(5), 1042–1055.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*(4), 713–726.

Buchsbaum, B. R., et al. (2005). Meta-analysis of neuroimaging studies of the Wisconsin Card-Sorting task and component processes. *Human Brain Mapping*, *25*(1), 35–45.

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.

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 117–126.

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.

Crosse, M. J., Foxe, J. J., & Molholm, S. (2019). Developmental recovery of impaired multisensory processing in autism and the cost of switching sensory modality. bioRxiv, 565333.

Crosse, M. J., et al. (2019). *Computational and systems neuroscience*, *Reconciling the cognitive and neural architecture of multisen-sory processing in the autistic brain*. 2019.

Cuppini, C., Magosso, E., & Ursino, M. (2009). A neural network model of semantic memory linking feature-based object representation and words. *Biosystems*, *96*(3), 195–205.

Cuppini, C., Magosso, E., & Ursino, M. (2011). Organization, maturation, and plasticity of multisensory integration: insights from computational modeling studies. *Frontiers in Psychology*, *2*, 77.

Cuppini, C., Magosso, E., & Ursino, M. (2013). Learning the lexical aspects of a second language at different proficiencies: A neural computational study. *Bilingualism*, *16*(2), 266–287.

Cuppini, C., Stein, B. E., & Rowland, B. A. (2018). Development of the mechanisms governing midbrain multisensory integration. *The Journal of Neuroscience*.

Cuppini, C., et al. (2010). An emergent model of multisensory integration in superior colliculus neurons. *Frontiers in Integrative Neuroscience*, *4*(6), 1–15.

Cuppini, C., et al. (2011). A computational study of multisensory maturation in the superior colliculus (SC). *Experimental Brain Research*, *213*(2–3), 341–349.

Cuppini, C., et al. (2012). Hebbian mechanisms help explain development of multisensory integration in the superior colliculus: a neural network model. *Biological Cybernetics*, *106*(11–12), 691–713.

Cuppini, C., et al. (2014). A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage*, *92*, 248–266.

Cuppini, C., et al. (2017a). A biologically inspired neurocomputational model for audiovisual integration and causal inference. *European Journal of Neuroscience*, *46*(9), 2481–2498.

Cuppini, C., et al. (2017b). A computational analysis of neural mechanisms underlying the maturation of multisensory speech integration in neurotypical children and those on the autism spectrum. *Frontiers in Human Neuroscience*, *11*(518), pp.

Dove, A., et al. (2000). Prefrontal cortex activation in task switching: an event-related fMRI study. *Cognitive Brain Research*, *9*(1), 103–109.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals cross-modal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23.

Foxe, J. J., Murphy, J. W., & De Sanctis, P. (2014). Throwing out the rules: anticipatory alpha-band oscillatory attention mechanisms during task-set reconfigurations. *European Journal of Neuroscience*, *39*(11), 1960–1972.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, *16*(5), pp.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285.

Gondan, M., et al. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin and Review*, *11*(2), 307–313.
Greenberg, A. S., et al. (2010). Control of spatial and feature-based attention in frontoparietal cortex. *Journal of Neuroscience*, *30*(43), 14330–14339.
Hairston, W. D., et al. (2008). Closing the mind's eye: deactivation of visual cortex related to auditory task difficulty. *Neuroreport*, *19*(2), 151–154.
Huang, S., et al. (2015). Multisensory competition is modulated by sensory pathway interactions with fronto-sensorimotor and default-mode network regions. *The Journal of Neuroscience*, *35*(24), 9064–9077.
Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports*, *9*(1), 1–10.
Jansen, B., & Rit, V. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, *73*(4), 357–366.
Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14*(89), 81.
Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, *18*(7), 1560–1574.
Koch, I., & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory & Cognition*, *34*(2), 433–444.
Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research*, *121*(1–2), 69–79.
Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Research. Cognitive Brain Research*, *16*(3), 468–478.
Liston, C., et al. (2006). Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. *Neuron*, *50*(4), 643–653.
Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press on Demand.
Magosso, E., Cuppini, C., & Bertini, C. (2017). Audiovisual rehabilitation in hemianopia: A model-based theoretical investigation. *Frontiers in Computational Neuroscience*, *11*(113), pp.
Magosso, E., Cuppini, C., & Ursino, M. (2012). A neural network model of ventriloquism effect and aftereffect. *Plos One*, *7*(8), Article e42503.
Magosso, E., et al. (2008). A theoretical study of multisensory integration in the superior colliculus by a neural network model. *Neural Networks*, *21*(6), 817–829.
Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology*, *68*(4), 1332–1344.
Mégevand, P., et al. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS One*, *8*(8), Article e71608.
Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Brain Research. Cognitive Brain Research*, *14*(1), 31–40.
Meredith, M. A., & Allman, B. L. (2015). Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. *European Journal of Neuroscience*, *41*(5), 686–698.
Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, *7*(10), 3215–3229.
Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279.
Miller, R. L., et al. (2015). Relative unisensory strength and timing predict their multisensory product. *Journal of Neuroscience*, *35*(13), 5213–5220.
Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing Research*, *258*(1–2), 72–79.
Navarra, J., et al. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Research. Cognitive Brain Research*, *25*(2), 499–507.
Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, *89*(1), 133.
Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *Journal of Neuroscience*, *33*(17), 7463–7474.
Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology*, *22*(15), 1391–1396.
Otto, T. U., & Mamassian, P. (2017). Multisensory decisions: the test of a race model, its logic, and power. *Multisensory Research*, *30*(1), 1–24.
Parise, C. V., et al. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration. *Multisensory Research*, *26*(3), 307–316.
Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*.
Raij, T., et al. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*, *31*(10), 1772–1782.
Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, *258*(1–2), 89–99.
Recanzone, G. H., Guard, D. C., & Phan, M. L. (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *Journal of Neurophysiology*, *83*(4), 2315–2331.
Reuter-Lorenz, P. A., Jha, A. P., & Rosenquist, J. N. (1996). What is inhibited in inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 367.
Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207.
Romei, V., et al. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. *The Journal of Neuroscience*, *27*(43), 11465–11472.
Rowland, B. A., & Stein, B. E. (2007). Multisensory integration produces an initial response enhancement. *Frontiers in Integrative Neuroscience*, *1*, 1–4.
Rowland, B. A., et al. (2007). Multisensory integration shortens physiological response latencies. *Journal of Neuroscience*, *27*(22), 5879–5884.
Shams, L., & Kim, R. (2010). Cross-modal influences on visual perception. *Physics of Life Reviews*, *7*(3), 269–284.
Shaw, L. H., et al. (2020). Operating in a multisensory context: Assessing the interplay between multisensory reaction time facilitation and inter-sensory task-switching effects. *Neuroscience*, *436*, 122–135.
Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *The American Journal of Psychology*, 669–679.
Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*(2), 330–336.
Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*(13), R519–R521.
Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, *9*(4), 255–266.
Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261.
Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network*, *4*, 259–284.
Ursino, M., Cuppini, C., & Magosso, E. (2010). A computational model of the lexical-semantic system based on a grounded cognition approach. *Frontiers in Psychology*, *1*(221).
Ursino, M., Cuppini, C., & Magosso, E. (2011). An integrated neural model of semantic memory, lexical retrieval and category formation, based on a distributed feature representation. *Cognitive Neurodynamics*, *5*(2), 183–207.
Ursino, M., Cuppini, C., & Magosso, E. (2014). Neurocomputational approaches to modelling multisensory integration in the brain: a review. *Neural Networks*, *60*, 141–165.
Ursino, M., Cuppini, C., & Magosso, E. (2017). Multisensory Bayesian inference depends on synapse maturation during training: theoretical analysis and neural modeling implementation. *Neural Computation*, *29*(3), 735–782.
Ursino, M., Magosso, E., & Cuppini, C. (2009b). Recognition of abstract objects via neural oscillators: Interaction among topological organization, associative memory and gamma band synchronization. *IEEE Transactions on Neural Networks*, *20*(2), 316–335.
Ursino, M., et al. (2009a). Multisensory integration in the superior colliculus: a neural network model. *Journal of Computational Neuroscience*, *26*(1), 55–73.
Ursino, M., et al. (2017). Development of a Bayesian estimator for audio-visual integration: A neurocomputational study. *Frontiers in Computational Neuroscience*, *11*(89).
Ursino, M., et al. (2018). A feature-based neurocomputational model of semantic memory. *Cognitive Neurodynamics*, *12*(6), 525–547.
Ursino, M., et al. (2019). Explaining the effect of likelihood manipulation and prior through a neural network of the audiovisual perception of space. *Multisensory Research*, *32*(2), 87–109.
Wallace, M. T., et al. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2), 252–258.
Wang, L., Yue, Z., & Chen, Q. (2012). Cross-modal nonspatial repetition inhibition. *Attention, Perception, & Psychophysics*, *74*(5), 867–878.
van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.

Weaver, S. M., et al. (2014). You can't always get what you want: The influence of unexpected task constraint on voluntary task switching. *Quarterly Journal of Experimental Psychology*, *67*(11), 2247–2259.

Wendling, F., et al. (2002). Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition. *European Journal of Neuroscience*, *15*(9), 1499–1508.

Williams, D. L., Goldstein, G., & Minshew, N. J. (2013). The modality shift experiment in adults and children with high functioning autism. *Journal of Autism and Developmental Disorders*, *43*(4), 794–806.

Wu, X., et al. (2019). Different visual and auditory latencies affect cross-modal non-spatial repetition inhibition. *Acta Psychologica*, *200*, Article 102940.

Wylie, G., Javitt, D., & Foxe, J. (2003a). Task switching: a high-density electrical mapping study. *Neuroimage*, *20*(4), 2322–2342.

Wylie, G., Javitt, D., & Foxe, J. J. (2003b). Cognitive control processes during an anticipated switch of task. *European Journal of Neuroscience*, *17*(3), 667–672.

Wylie, G. R., Javitt, D. C., & Foxe, J. J. (2004). Don't think of a white bear: An fMRI investigation of the effects of sequential instructional sets on cortical activity in a task-switching paradigm. *Human Brain Mapping*, *21*(4), 279–297.

Wylie, G. R., et al. (2009). Distinct neurophysiological mechanisms mediate mixing costs and switch costs. *Journal of Cognitive Neuroscience*, *21*(1), 105–118.

Yu, L., et al. (2013). Multisensory plasticity in adulthood: cross-modal experience enhances neuronal excitability and exposes silent inputs. *Journal of Neurophysiology*, *109*(2), 464–474.