Appendix: Mathematical description of the model

In the following, each neuron will be denoted with a superscript, c, referred to a specific cortical area ($c = a$ or v or m, for the auditory, visual, or multisensory region, respectively), and a subscript, j, which indicates the spatial position within that area. $u(t)$ and $y(t)$ are used to represent the net input and output of a given neuron at time t. Thus, $y_j^c(t)$ represents the output of a unit at position j, belonging to the area c, described by the following differential equation:

$$\tau^c \frac{dy_j^c(t)}{dt} = -y_j^c(t) + F\left(u_j^c(t)\right) \tag{1}$$

where τ^c is the time constant of neurons belonging to the area c, and $F(u)$ represents a sigmoidal relationship:

$$F(u_j^c) = \frac{1}{1 + e^{-s(u_j^c - \theta)}} \tag{2}$$

s and θ are parameters which establish the slope and the central position of the sigmoidal relationship, respectively. The saturation value is set at 1, i.e., all activities are normalized to the maximum.

For the sake of simplicity, in this work neurons belonging to the three regions differ only for the time constants, chosen to mimic a quicker sensory processing for stimuli in the auditory region compared to visual stimuli.

The net input that reaches a neuron (i.e., the quantity $u_j^c(t)$ in Eq. 1) is the sum of two terms: a *within-region component* (say $l_j^c(t)$), due to the contribution of lateral synapses from other neurons in the same area, and a component coming from *extra-area sources* (say $i_j^c(t)$). Hence, we have

$$u_j^c(t) = l_j^c(t) + i_j^c(t); \tag{3}$$

To simulate the lateral input, $l_j^c(t)$, neurons within each area interact via excitatory and inhibitory lateral synapses: each neuron excites (and is excited by) its proximal neurons, and inhibits (and is inhibited by) more distal neurons. Thus, the input that a neuron receives from other elements of the same area is defined as:

$$l_j^c(t) = \sum_k L_{jk}^c \cdot y_k^c(t);$$

(4)

where L_{jk}^c is the strength of the lateral synapse from a presynaptic neuron at position k to a postsynaptic neuron at position j both belonging to the same region c, and $y_k^c(t)$ is the activity of the presynaptic neuron at position k. These synapses are symmetrical and arranged according to a "Mexican hat" distribution (a central excitatory zone surrounded by an inhibitory annulus):

$$L_{jk}^c = \begin{cases} L_{ex0}^c \cdot e^{-\frac{(d_{jk})^2}{2(\sigma_{ex}^c)^2}} - L_{in0}^c \cdot e^{-\frac{(d_{jk})^2}{2(\sigma_{in}^c)^2}} & \text{if } d_{jk} \neq 0; \\ \\ 0 & \text{if } d_{jk} = 0; \end{cases}$$

(5)

In this equation, L_{ex0} and σ_{ex} define the excitatory Gaussian function, while L_{in0} and σ_{in} the inhibitory one, and d_{jk} represents the distance between the pre-synaptic and post-synaptic neurons in the same area. To avoid undesired border effects, we use a circular structure to realize these synapses so that every neuron in each area receives the same number of side connections. This is obtained assuming the following expression for the distance:

$$d_{jk} = \begin{cases} |j-k| & \text{if } |j-k| \leq N/2 \\ N - |j-k| & \text{if } |j-k| > N/2 \end{cases}$$

(6)

In this work, we assume that both unisensory areas have the same pattern of lateral synapses, in order to limit the number of hypotheses in building the model.

The external component of the input, $i_j^c(t)$, has a different expression for the unisensory areas ($c = a$, v) and for the multisensory one ($c = m$).

The input to each *unisensory area* includes: a sensory stimulus from the external world (say $e_j^c(t)$), a cross-modal term coming from the other unisensory area (say $c_j^c(t)$), and a *noise component*, n_j^c, realized by a standard uniform distribution on an interval $[-n_{max} \; +n_{max}]$, where n_{max} is equal to the 40% of the strength of the external stimulus for each sensory modality. Hence

$$i_j^c(t) = e_j^c(t) + c_j^c(t) + n_j^c \qquad c = a, v \qquad (7)$$

The first term in (7) is simulated by means of a spatial Gaussian function, to reproduce the uncertainty in the detection of external stimuli. Assuming a stimulus of sensory modality c ($c = a$ or v) presented in the position p^c, the consequent input to the network can be written as

$$e_j^c(t) = E_0^c \cdot e^{-\frac{\left(d_j^c\right)^2}{2\left(\sigma^c\right)^2}} \qquad (8)$$

where E_0^c represents the strength of the stimulus, d_j^c is the distance between the neuron at position j and the stimulus at position p^c, and σ^c defines the degree of uncertainty in sensory detection (which establishes the overall number of elements in region c, activated by the same external stimulus). As previously described for the lateral synapses, to avoid undesired border effects, also the external inputs are implemented as having a circular structure; hence, the distance d_j is defined as:

$$d_j^c = \begin{cases} \left|j - p^c\right| & \text{if } \left|j - p^c\right| \leq N/2 \\ N - \left|j - p^c\right| & \text{if } \left|j - p^c\right| > N/2 \end{cases} \qquad (9)$$

The cross-modal input, $c_j^c(t)$, is obtained assuming that each unisensory neuron receives an excitation from the neurons processing the other modality. Hence

$$c_j^a(t) = \sum_{k=1}^{N} W_{jk}^{av} \cdot y_k^v(t)$$

$$c_j^v(t) = \sum_{k=1}^{N} W_{jk}^{va} \cdot y_k^a(t)$$

(10)

The weights of these excitatory connections, W_{ij}^{av}, W_{ij}^{va}, are described by a gaussian function, i.e.

$$W_{jk}^{cd} = W_0^{cd} \cdot e^{-\frac{(d_{jk})^2}{2 \cdot (\sigma^{cd})^2}} \qquad cd = av \text{ or } va$$

(11)

where W_0 is the highest level of synaptic efficacy, d_{jk} is the distance between the neuron at position j in the post-synaptic unisensory region and the unisensory neuron at position k in the pre-synaptic area. σ^{cd} defines the width of the cross-modal synapses. The patterns of the cross-modal synapses are symmetrical ($W_0^{av} = W_0^{va}$ and $\sigma^{av} = \sigma^{va}$).

Finally, the external component of the input to the multisensory neurons is only due to the feedforward connections from the two unisensory areas. Hence

$$i_j^m(t) = \sum_{k=1}^{N} W_{jk}^{ma} \cdot y_k^a(t) + \sum_{k=1}^{N} W_{jk}^{mv} \cdot y_k^v(t)$$

(12)

where W_{jk}^{ma} and W_{jk}^{mv} are the synapses linking the presynaptic neuron at position k in the unisensory area (auditory and visual, respectively) to the jth neuron in the multisensory area. The effectiveness of these synapses is described by a Gaussian shape, i.e

$$W_{jk}^{mc} = W_0^{mc} \cdot e^{-\frac{(d_{jk})^2}{2 \cdot (\sigma^{mc})^2}} \qquad c = a, v$$

(13)

where W_0^{mc} is the highest level of synaptic efficacy, d_{jk} is the distance between the multisensory neuron at position j and the unisensory neuron at position k. σ^{mc} defines the width of the feedforward synapses.

In the model for simplicity, we set the feedforward synapses identical for the two modalities

Parameter assignment

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

External inputs – Physiological evidences show that in the brain the visual system presents better spatial resolution than the auditory one (see (Recanzone, 2000; 2009)). This is mimicked by setting σ^a $> \sigma^v$ In particular, the value of σ^v is assigned to have a very acute visual perception (with a few degree uncertainty). Then, the value for σ^a is set, to have a large ratio σ^a / σ^v, according to our previous computational studies (Magosso *et al.*, 2012; Cuppini *et al.*, 2014).

It is worth noting that this is the only difference between the two unisensory areas (auditory and visual); all other parameters for the unisensory neurons are assumed equal, regardless of their respective area, to reduce the number of assumptions.

The strength of the external visual and auditory stimuli (parameters E_0^v and E_0^a) is chosen so that the overall input elicits a response in the upper portion of the linear part of the sigmoidal static characteristic (i.e., a little below saturation).

Parameters of individual neurons – The central abscissa, θ, is assigned to have negligible neuron activity in basal conditions (i.e., when the input is zero). The slope of the sigmoidal relationship, *s*, is assigned to have a smooth transition from silence to saturation in response to external stimuli. The time constants agrees with values (a few ms) normally used in deterministic mean-field equations (Treves, 1993; Ben-Yishai *et al.*, 1995). They are set so that the auditory processing is faster than the visual one,

$\tau^a < \tau^v$, in accordance to experimental evidences of auditory cortical neurons presenting shorter latencies than neurons in the visual cortex (Maunsell & Gibson, 1992; Recanzone, 2000).

Parameters of the lateral connections – Parameters which establish the width and the strength of lateral synapses in every areas (i.e., L_{ex0}, L_{in0}, σ_{ex} and σ_{in}) are assigned to simultaneously satisfy several criteria: (1) inhibition must be strong enough to warrant competition between two stimuli in the same area, (2) the balance between excitation and inhibition must avoid instability, i.e., an uncontrolled excitation which propagated to the overall area, and (3) the width of inhibitory synapses is large in the unisensory regions, to realize a stronger competition between two inputs of the same modality. Conversely, (4) the width of inhibitory synapses in the multisensory region is smaller, in agreement with the possibility to have the coexistence of two peaks of activity in case of external unrelated stimuli, even at a few degree distance.

Parameters of the inter-area synapses – The strength and width of the feedforward synapses (1) provide an input to the multisensory neurons close to the central portion of the sigmoidal I/O relationship, when the network is stimulated by a strong unisensory external stimulus, but in the upper part close to saturation when stimulated by two concordant multisensory stimuli. This corresponds to the classic principle of multisensory integration (enhancement, inverse effectiveness). In particular, with this choice we have super-additivity in case of weak unisensory stimuli, and sub-additivity in case of stronger stimuli. The direct synapses among the unisensory regions are assigned so that: (2) these synapses can affect the response of modality-specific neurons when these elements are near or just above the activation threshold; (3) the report of unity and auditory localization bias are in satisfactory agreement with behavioral data. However these synapses are maintained sufficiently low so that (4) an external stimulus in one sensory modality does not induce a phantom activity in the other modality-specific area.

A list of all model parameters is reported in Table S1.

Table S1 – paramenters value

Neurons	Input
N = 180	$E_0^v = 27$
$\theta = 20$	$E_0^a = 28$
$s = 0.3$	$\sigma^v = 4$
$\tau = 1\text{ms}$	$\sigma^a = 32$
$\tau^v = 15\text{ms}$	$n_{max} = 40\%$
$\tau^a = 3\text{ms}$	**Lateral Synapses Multisensory Area**
Lateral Synapses Unisensory Areas (same parameters in the two areas)	$L_{ex0} = 3$
$L_{ex0} = 5$	$\sigma_{ex} = 2$
$\sigma_{ex} = 3$	$L_{in0} = 2.6$
$L_{in0} = 4$	$\sigma_{in} = 10$
$\sigma_{in} = 120$	**Feedforward Synapses**
Cross-modal Synapses	$W_0^{ma} = W_0^{mv} = 18$
$W_0^{av} = W_0^{va} = 1.4$	$\sigma^{ma} = \sigma^{mv} = 0.5$
$\sigma^{av} = \sigma^{va} = 5$	

FURTHER SENSITIVITY ANALYSES

A) $L_{in}0 = 3$ (Rohe and Noppeney, 2015)
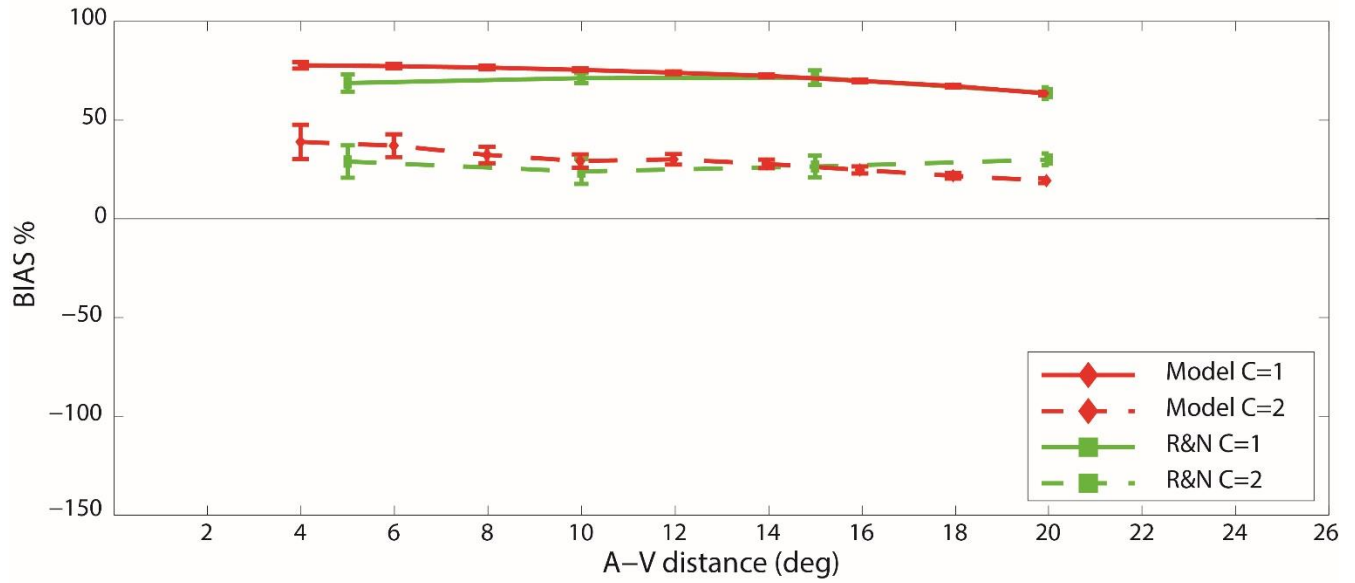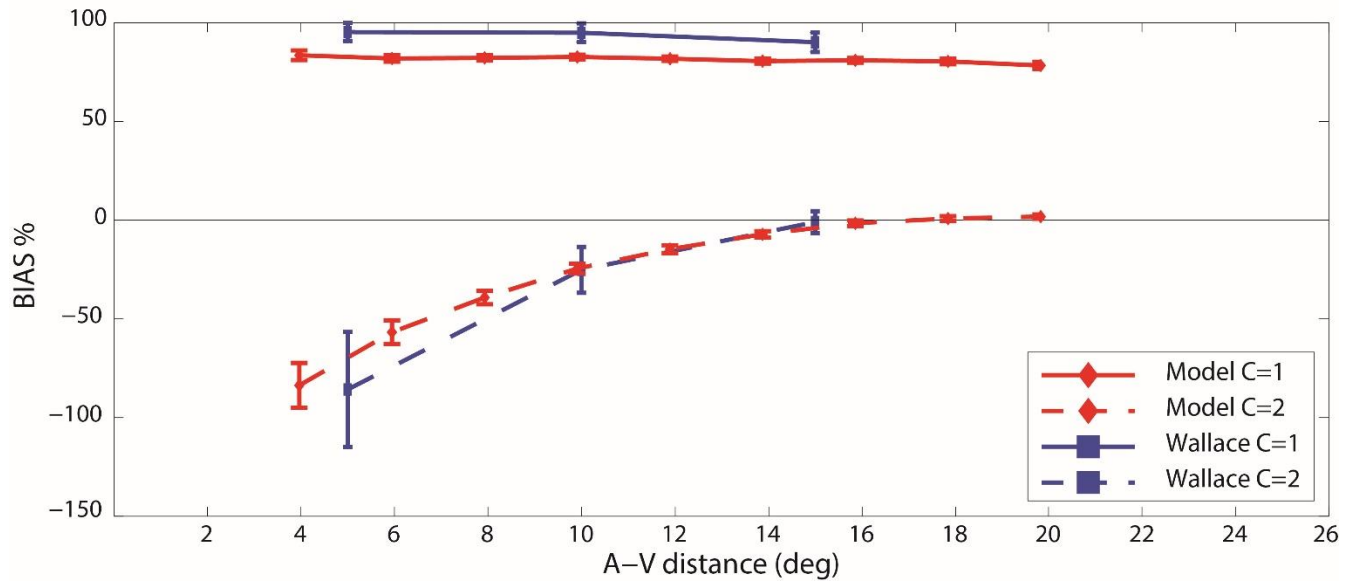

B) $L_{in}0 = 6$ (Wallace et al., 2004)

Figure S1 – Sensitivity Analysis on the competitive mechanism in the unisensory regions for the auditory perception bias, evaluated separately when the network identifies a common cause (C=1, red solid line) or different causes (C=2, red dashed line). The figure shows the effect of modifying the strength of the lateral competition mechanism in the unisensory areas, on the auditory localization bias in case of common cause and independent cause evaluation, at different A-V distances. The strength of lateral synapses (L^a_{ex0}, L^v_{ex0}, L^a_{in0} and L^v_{in0}) has been varied, by maintaining a constant ratio between excitation and inhibition (i.e. Lin0/Lex0 constant). A) A weak competition in the auditory and visual regions is suitable to reproduce the results of Rohe and Noppeney (2015) of a positive bias in case of independent sources (C=2). Conversely, B) a stronger inhibition in the unisensory regions is able to explain the strong negative bias identified by Wallace and colleagues (2004) in case of separate cause evaluation.
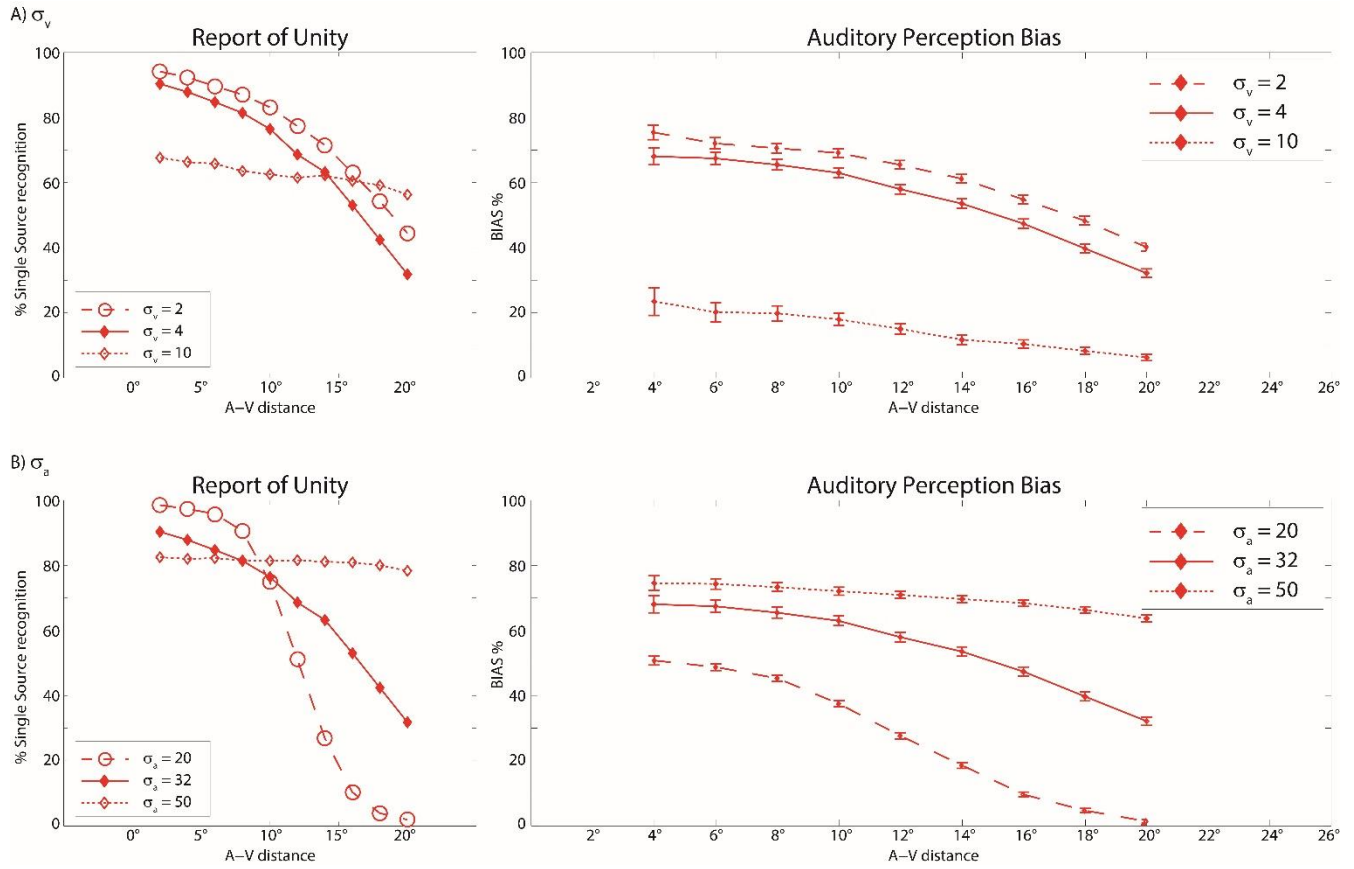
Figure S2 – Effect of the A) visual and B) auditory input SD. Decreasing the SD of the two stimuli (i.e., using more accurate stimuli) makes the report of unit sharper (i.e., closer to one at small distances, closer to zero at larger distances). Conversely, the auditory perception bias depends on the ratio between the accuracy of the two stimuli: stimuli with larger difference in their accuracy are associated with a very strong auditory bias, whereas stimuli with more proximal accuracy are associated with smaller auditory bias.
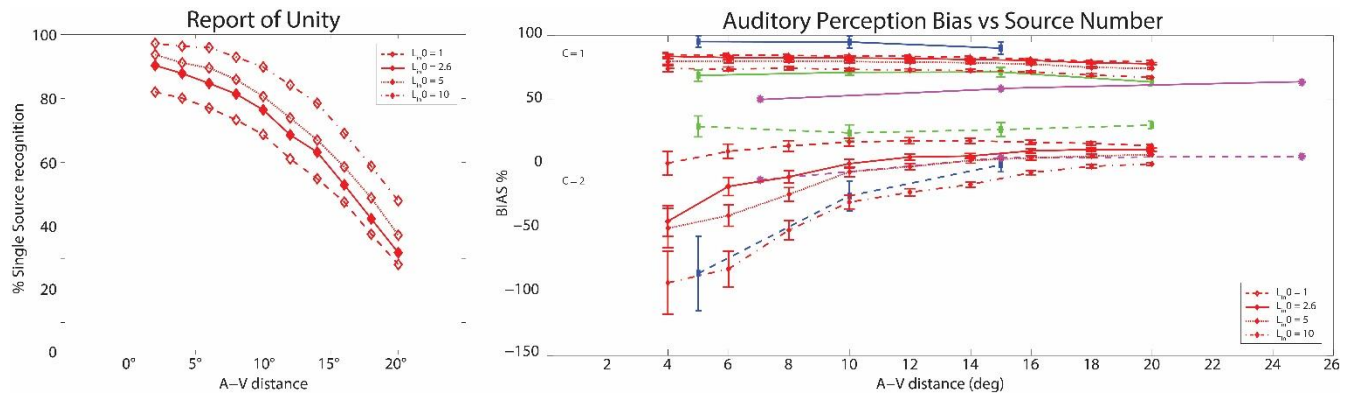

Figure S3- Sensitivity Analysis on the mechanisms operating in the multisensory area. The figure shows the effect of changing the strength of the lateral competition in the multisensory area, on the Report of Unity and the auditory localization bias in case of C=1 and C=2, at different A-V distances. The strength of lateral synapses (L_{ex0} and L_{in0}) has been varied, by maintaining a constant ratio between excitation and inhibition (i.e. Lin0/Lex0 constant). In all panels, the basal condition is displayed with a continuous line. In terms of the ability to discriminate stimuli at different spatial positions, the competition in the multisensory area has the opposite effect compared with the same mechanism present in the unisensory regions: a stronger inhibition leads to a greater likelihood of the perception of a common cause, even with stimuli presented at great distances. In this case, a peak of evoked activity in the multisensory region inhibits the activation of nearby elements in the same area; this generates the perception of a common cause for multisensory stimuli even at great distances.
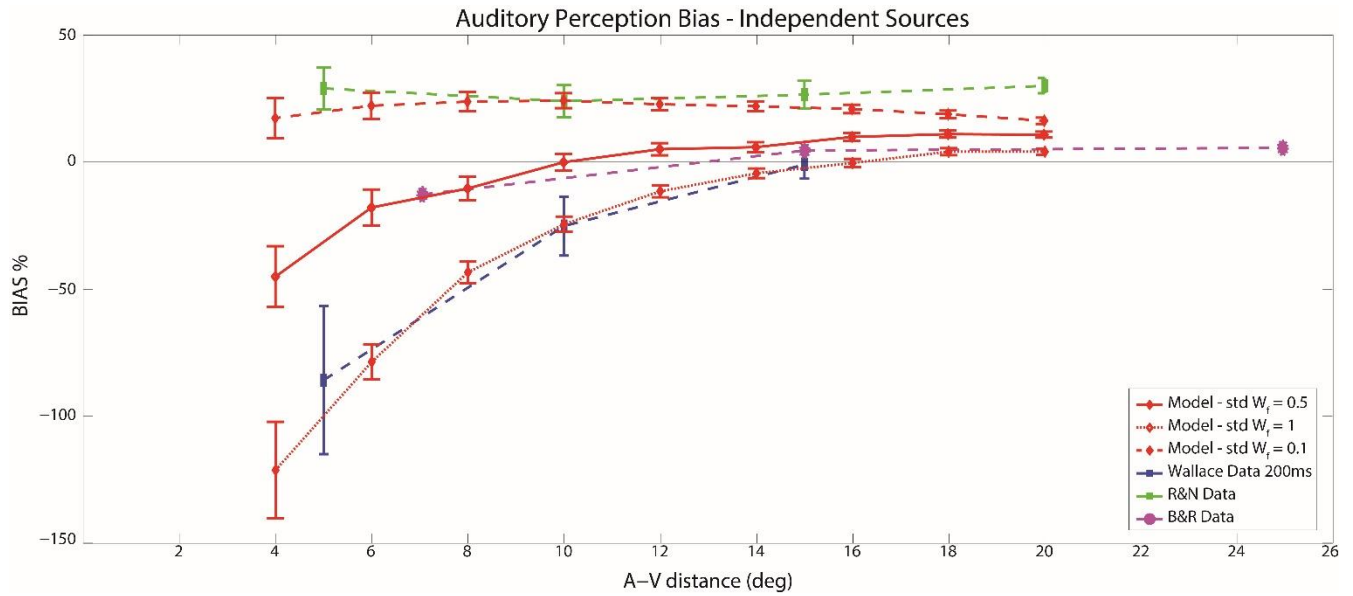
Figure S4 – Effect of the feedforward connectivity on the auditory perception bias in case of independent causes. The figure shows a possible explanation for differences in behavioral results present in the literature: the model is able to reproduce different experimental data by varying the feedforward connectivity. A neural architecture characterized by few feedforward connections (simulated with a small standard deviation, $\sigma = 0.1$, dashed red line) presents a positive bias also in case of independent input sources evaluation, similar to the results of Rohe and Noppeney (2015). Increasing the number of connections from the unisensory regions to the multisensory area, the model reproduces behavioral results of Bertelson and Radeau (1981), for $\sigma = 0.5$, and Wallace et al., (2004), for $\sigma = 1$.