ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

A biologically inspired neurocomputational model for audiovisual integration and causal inference

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Cuppini, C., Shams, L., Magosso, E., Ursino, M. (2017). A biologically inspired neurocomputational model for audiovisual integration and causal inference. EUROPEAN JOURNAL OF NEUROSCIENCE, 46(9), 2481-2498 [10.1111/ejn.13725].

Availability:
This version is available at: https://hdl.handle.net/11585/614817 since: 2018-01-09

Published:
DOI: http://doi.org/10.1111/ejn.13725

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

12 October 2024

This is the final peer-reviewed accepted manuscript of:

Cuppini, C., Shams, L., Magosso, E. and Ursino, M. (2017), ***A biologically inspired neurocomputational model for audiovisual integration and causal inference****. Eur J Neurosci, 46: 2481-2498*

The final published version is available online at: https://doi.org/10.1111/ejn.13725

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

DR. CRISTIANO CUPPINI (Orcid ID : 0000-0002-6529-2599)

PROF. MAURO URSINO (Orcid ID : 0000-0002-0911-0308)

Author Manuscript

Article type : Research Report

Journal Section: Cognitive Neuroscience

A biologically inspired neurocomputational model for audio-visual integration and causal inference

Cristiano Cuppini[1*], Ladan Shams[2], Elisa Magosso[1], Mauro Ursino[1]

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ejn.13725

This article is protected by copyright. All rights reserved

[1]Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy

[2]Department of Psychology, Department of BioEngineering, Interdepartmental Neuroscience Program, University of California, Los Angeles

Corresponding author:
Cristiano Cuppini
Department of Electrical, Electronic and Information Engineering, University of Bologna
Viale Risorgimento 2, I40136, Bologna, Italy
Email: cristiano.cuppini@unibo.it

Running title: Neurocomputational analysis of causal inference

Number of pages: 40
Number of figures: 13
Number of tables: 1
Number of equations: 0
Words in the manuscript: 10560
Words in the Abstract: 248
Keywords: multisensory integration; neural network; integrative mechanisms; ventriloquism effect; spatial sensory processing

Abstract

Recently, experimental and theoretical research has focused on the brain's abilities to extract information from a noisy sensory environment and how cross-modal inputs are processed to solve the causal inference problem to provide the best estimate of external events. Despite the empirical evidence suggesting that the nervous system uses a statistically optimal and probabilistic approach in addressing these problems, little is known about the brain's architecture needed to implement these computations.

The aim of this work is to realize a mathematical model, based on physiologically plausible hypotheses, to analyze the neural mechanisms underlying multisensory perception and causal inference. The model consists of three layers topologically organized: two encode auditory and

Author Manuscript

This article is protected by copyright. All rights reserved

visual stimuli, separately, and are reciprocally connected via excitatory synapses and send excitatory connections to the third downstream layer. This synaptic organization realizes two mechanisms of cross-modal interactions: the first is responsible for the sensory representation of the external stimuli while the second solves the causal inference problem.

We tested the network by comparing its results to behavioral data reported in the literature. Among others, the network can account for the ventriloquism illusion, the pattern of sensory bias and the percept of unity as a function of the spatial auditory-visual distance, and the dependence of the auditory error on the causal inference. Finally, simulations results are consistent with probability matching as the perceptual strategy used in auditory-visual spatial localization tasks, agreeing with the behavioral data. The model makes untested predictions that can be investigated in future behavioral experiments.

Introduction

Perception of objects in the external world requires the integration of information from different modalities, for example auditory and visual. Several recent behavioral (Ernst & Banks, 2002; Battaglia *et al.*, 2003; Alais & Burr, 2004; Wallace *et al.*, 2004; Shams *et al.*, 2005; Hillis *et al.*, 2006; Wozny *et al.*, 2008) and theoretical studies (Pouget *et al.*, 2003; Ma *et al.*, 2006; Shams & Beierholm, 2010; Ma & Rahamati, 2013; Pouget *et al.*, 2013) suggest that the brain performs this integration in a Bayesian way, i.e., it tries to exploit the different multisensory signals to minimize the error in perceptual estimates (the error on the spatial location of the stimuli, for instance). Two factors are involved in achieving this estimate. First, the observer must infer whether the two stimuli (e.g., a sound and a light) originate from the same event (i.e., they have a common cause) or stem from different sources (i.e., they are produced by independent causes). This problem is referred to as "the causal inference problem": its solution is crucially dependent on the similarity of the different sensory features (in cases of simple stimuli, the spatial and temporal proximity) and previous knowledge/bias or expectations. Causal inference only recently received sufficient attention in multisensory neuroscience literature (Shams *et al.*, 2005; Körding *et al.*, 2007; Wozny *et al.*, 2008; Shams & Beierholm, 2010; Wozny *et al.*, 2010; Cuppini *et al.*, 2012; Ma & Rahamati, 2013).

Second, in cases when a single cause is inferred, the two stimuli must be integrated to realize a coherent percept; conversely, in case of two independent causes, they should be treated separately. A large body of results demonstrates that, in cases of a single cause, the observer integrates the

This article is protected by copyright. All rights reserved

Author Manuscript

signals based on their sensory reliability (Alais & Burr, 2004; Morgan *et al.*, 2008; Fetsch *et al.*, 2012), giving more weight to the more reliable cue. An epiphenomenon of this integration is the occurrence of illusory phenomena in certain conditions in which a discrepancy exists between the two sensory signals, and yet the more reliable input attracts the other, as in the spatial ventriloquism (Bertelson & Radeau, 1981; Hairston *et al.*, 2003; Wallace *et al.*, 2004; Rohe & Noppeney, 2015b) or in the temporal fission phenomenon (Shams *et al.*, 2000; 2002; Andersen *et al.*, 2004; Shams *et al.*, 2005).

Various models, based on Bayesian inference, have appeared recently, offering a theoretical foundation for multisensory integration. Some of these models provide results in good agreement with behavioral data, in a variety of multisensory tasks (Shams *et al.*, 2005; Körding *et al.*, 2007; Wozny *et al.*, 2008; 2010; Samad *et al.*, 2015). However, these models make direct use of probabilities to compute the estimates, without implementing a biologically-inspired neural representation.

A related issue, yet insufficiently understood, is how neural circuits can implement the computation of these probabilities, which are necessary to infer the correct estimate. This question can be approached by searching for a biologically inspired neural network, which can produce the same estimate as the Bayesian model (or, alternatively, can produce estimates in good agreement with behavioral data).

A possible solution to this problem is provided by the so-called "neural population coding" (Pouget *et al.*, 2003; Ma *et al.*, 2006; Pouget *et al.*, 2013). In this scenario the activity of a population of neurons encodes the probabilities of the individual realization of a given variable (for instance, the position of the auditory or visual signal) in a trial-by-trial basis. The positions are then estimated, from the overall population activity, using some kind of metrics (such as the maximum activity or the barycenter).

Following these ideas, a few models have been proposed in recent years, which exploit some biologically inspired representations (Ma & Rahamati, 2013; Yamashita et al., 2013; Parise & Ernst, 2013; Zhang et al., 2016) and in some cases explicitly consider the causal inference problem (Ma & Rahamati, 2013; Yamashita et al., 2013).

In a series of recent studies (Magosso *et al.*, 2012; Magosso *et al.*, 2013; Cuppini *et al.*, 2014), we constructed a biologically plausible neural network, which incorporates two chains of unisensory neurons (one auditory and one visual) linked via cross-modal synapses. With this model, we were able to demonstrate that illusory phenomena crucially depend on the cross-modal synapse weights, which implement a prior on the co-occurrence of the stimuli. Various recent experimental data support this model structure, showing that primary areas in the cortex (such as the V1 and A1),

Author Manuscript

This article is protected by copyright. All rights reserved

traditionally deemed as purely unisensory, can exchange reciprocal information and reciprocally influence the other's activities (Ghazanfar & Schroeder, 2006; Musacchia & Schroeder, 2009; Recanzone, 2009; Ursino *et al.*, 2014).

In our previous model, however, the causal inference problem was not explicitly considered and the circuit provided no estimate on the number of sources. The aim of this work is to improve significantly our previous studies, by proposing a network circuit that not only estimates the individual positions, but can also infer the number of external sources. This is obtained by adding a third layer of multisensory neurons (which can mimic activities in higher hierarchical regions of the cortex) that receive inputs from the two unisensory layers. The aim of the multimodal layer is to integrate audio-visual inputs, according to the classic laws of multisensory integration (basically the spatial proximity, the temporal proximity, the enhancement, and the inverse effectiveness), and to compare the resulting excitation with a given threshold, to infer whether the stimuli originate from a common cause or derive from independent causes. It is worth noting that this multi-layer structure is in part supported by the hierarchical nature of multisensory integration as recently proposed by Rohe and Noppeney on the basis of neuroimaging data (Rohe & Noppeney, 2015a). These authors suggest that at the early stages of the hierarchy, in auditory and visual areas, locations are represented without taking into account the underlying causal structure. The estimation that the two signals derive from a common cause is performed only at a successive stage, in the anterior intraparietal sulcus.

Here, we first present the structure of the model, with a focus on the main mechanisms and their putative role (the mathematical description is provided in the Supplementary Material). Next, the results of simulations (with a basal set of parameters) are presented and compared with behavioral data, concerning the causal inference and the auditory localization bias. Finally, a sensitivity analysis on model mechanisms is performed, in order to shed light on their role in multisensory causal inference and integration.

Method

The model's architecture (Figure 1) is based on previous networks realized to study audio-visual multisensory processes, such as perceptual illusions and speech integration (see (Magosso *et al.*, 2012; Cuppini *et al.*, 2014; Ursino *et al.*, 2017). In the following, we will describe the mechanisms implemented in the network and its most important emerging behaviors. We will then present the

Author Manuscript

This article is protected by copyright. All rights reserved

structure of the model and the simulations realized to test its abilities. All mathematical equations and parameter assignment criteria are given in the Supplementary Material.

Main model mechanisms

The network consists of two unisensory regions that process noisy external auditory and visual stimuli, respectively, and are reciprocally linked by means of cross-modal excitatory connections. Neurons in these regions are topologically organized, i.e., proximal neurons code for proximal spatial positions. These areas simulate the level of sensory processing performed in the unisensory cortical regions of the brain, and are responsible for inferring the spatial location of the sensory stimuli. In the model, the perceived position of the external stimuli is obtained by computing the barycenter of the activities elicited in the visual and auditory areas, respectively. Due to the presence of cross-modal projections between these two regions, the inferred spatial localization of the auditory or visual inputs is affected by the concurrent presentation of the stimulus in the other sensory modality, even if the two events are processed separately in the two unisensory regions.

While the presence of a topographic organization is well documented in the primary visual areas, this has not been observed in the acoustic areas. Therefore, the acoustic area must be considered as functionally equivalent to several stages of processing in the auditory cortex.

Information regarding the stimuli spatial configuration, extracted by these regions, is sent to a multisensory area (simulating an association cortex, for example, the anterior intraparietal sulcus, as shown by Rohe and Noppeney (Rohe & Noppeney, 2015a)). The role of this region is to solve the causal inference problem: are the stimuli produced by the same event or do they belong to different input sources?

To answer this question, the activity elicited in the multisensory region is compared with a threshold (the "Detection Threshold"). The number of distinct peaks of activity, in the multisensory region, above this threshold identifies the number of distinct input sources inferred by the model. Stimuli placed in proximal positions (i.e., likely caused by the same event) excite proximal neurons in the multisensory region, producing a single peak of activity above the threshold. Conversely, stimuli from different spatial positions (i.e., likely generated by different events) stimulate distant multisensory neurons, eliciting multiple peaks above the threshold in the multisensory area.

In this work, we used the barycenter method to identify the location of a stimulus. We found three main methods in the literature to extract the output from a population of neurons: the barycenter, the maximum activity, and the vector population decoder. We tried all of them within our model. The barycenter method and the vector population decoder produce practically the same

Author Manuscript

This article is protected by copyright. All rights reserved

results. The maximum activity method has the disadvantage of providing a discrete representation of space, and is more affected by noise. For this reason, we preferred the barycenter method.

From a biological point of view, both the barycenter and threshold effects can be easily implemented with a downstream layer of neurons.

Accordingly, this network presents two different levels of multisensory processing: the first, at the level of unisensory regions, makes a judgement about the spatial position of the external stimuli (some authors refer to this process as an "implicit causal inference", (Rohe & Noppeney, 2015a)); the second, performed in the multisensory region, is responsible for the solution of the causal inference problem (the "explicit causal inference" in (Rohe & Noppeney, 2015a)).

A further important mechanism in the model consists of competition/cooperation between elements in the same area. This is achieved via intra-area (lateral) synapses linking elements belonging to the same region, realized through a Mexican Hat disposition, so that elements sensitive to proximal portions of the external world excite one-another, and elements sensitive to different portions of the space are reciprocally inhibited. This synaptic arrangement concurs to identify the minimal distance between two activities in the same area that the network can separate, and thus are identified as produced by different events.

To summarize the role of the main mechanisms delineated above:

i) The two external inputs (auditory and visual) separately excite the two unisensory areas.

ii) The cross-modal synapses between unisensory regions modify the spatial perception of the sensory inputs. In cases of proximal stimuli, which are usually perceived as originating from a common cause, the two positions are reciprocally attracted, thus generating typical perceptual illusions (such as the ventriloquism). In cases of distant stimuli, which are usually perceived as coming from distinct input sources, these cross-modal synapses have a less important role, and intra-area inhibition becomes the dominant mechanism.

iii) The feedforward synapses from the unisensory input regions realize a classic multisensory integration, i.e., the enhancement of the activities in spatial register. This is used to encode information on the mutual spatial coincidence of the cross-modal stimuli, and the likelihood that two stimuli were generated by a common source. Indeed, when two multisensory stimuli fall inside the receptive fields (RFs) of the same multisensory neurons, the multisensory area integrates the information and presents a unique peak of activity, identifying a single input source.

iv) The inhibitory lateral synapses implement a competitive mechanism, which allows the survival of the stronger stimuli only, while spurious or negligible stimuli are suppressed. This has two fundamental functions: it favors a spatial shift in the unisensory areas (where less reliable stimuli are shifted in the direction of the more reliable ones), and it engenders the effective

Author Manuscript

This article is protected by copyright. All rights reserved

elimination of unimportant sources in the multisensory area, where the readout of causal inference is effectively realized.

Model structure

Each area consists of an array of 180 elements, topographically organized, so that each element is responsive to a specific portion of the external space. We assume a distance of 1° between adjacent elements. Each neural unit in every region is described by a sigmoidal I/O relationship and a first order dynamics (chosen to mimic a quicker sensory processing for stimuli in the auditory region compared to visual stimuli), and it is linked with units belonging to the same region through lateral synapses arranged with a Mexican Hat disposition. Moreover, elements in the unisensory regions are reciprocally connected with elements of the other unisensory area via excitatory inter-area synapses (cross-modal synapses). The presence of a sigmoidal relationship, with an "activation threshold" is important. In fact, many neurons (especially in the auditory region) are silent but close to this threshold and can be easily excited as a consequence of noise or cross-modal influences.

Accordingly, the net input reaching a neuron in the unisensory regions is the sum of three components: an external input, a multisensory input from neurons in the other modality (via cross-modal synapses W^{av}, W^{va}), and a lateral input coming from other neurons in the same unisensory area (via lateral synapses L^a and L^v). Moreover, to mimic the variability of sensory stimuli in a real environment, we added a noisy component targeting every element in the unisensory regions.

It is worth noting that we mimicked auditory localization in the same way as in the visual area. This is a strong simplification, since such topological organization is not present in the auditory regions in the brain. This aspect is further commented in the Discussion, where lines for future improvements are delineated.

Finally, units in the multisensory region receive inputs from elements of the unisensory layers that are sensitive to the same portion of the space, through excitatory feedforward synapses (W^{mv}, W^{ma}), and a lateral input, generated by the lateral synapses linking elements in the multisensory region (L^m).

External inputs

The visual and auditory inputs are described with a Gaussian function to mimic spatially localized external stimuli, filtered by neurons' receptive fields (RFs). The central point of the Gaussian function corresponds to the application point of the stimulus in the external world (p^a and p^v, for the auditory and visual stimuli, respectively); the standard deviation of the Gaussian function

Author Manuscript

This article is protected by copyright. All rights reserved

(σ^a, σ^v) reflects the width of neurons' RFs and the reliability of the external input. These parameters mimic the different spatial acuity of the auditory and visual sensory modalities.

It is worth noting that these inputs summarize two effects together: the SD of the present inputs and the receptive field (RF) of the unisensory neuron. As shown in Ursino et al. (Ursino *et al.*, 2017) both terms influence the perceived neural input (that should be considered as the convolution of the input and the RF). The two terms have been condensed for simplicity, but can be analyzed separately in future model implementations (as in (Ursino *et al.*, 2017)).

As for the temporal properties of the external stimuli, for simplicity in this work we chose synchronized auditory and visual stimuli, kept constant throughout the simulations (except for the analysis of the temporal window for the multisensory integration, see below in the Method section).

Cross-modal terms

The multisensory input is computed assuming that neurons of the unisensory areas covering proximal portions of the external space are reciprocally connected via excitatory synapses. The connectivity is symmetrical, and it is realized by using a Gaussian function.

Feedforward terms

Elements in the multisensory region receive feedforward excitatory connections from unisensory neurons coding for proximal portions of the external world. Again, we used a Gaussian function, equal for the two modalities.

The lateral connections

The lateral input originates from connections within the same layer. These connections include both excitatory and inhibitory lateral synapses, which are arranged with a classic Mexican Hat disposition (a central excitatory zone surrounded by an inhibitory annulus). Therefore, each neuron excites (and is excited by) its proximal neurons, and inhibits (and is inhibited by) more distal neurons. Hence, activities of neurons belonging to the same region and stimulated by distal stimuli tend to suppress reciprocally (i.e., they interact via a competitive mechanism). For simplicity, in the network we implemented the same lateral connectivity among elements in the unisensory regions, but we used a different connectivity in the multisensory layer to improve solution of the causal inference problem.

Finally, we wish to stress that, in the present model, the two sensory modalities differ only for the respective receptive fields (RF) and time constants, specifically the visual region presents smaller RF but greater time constant than the auditory one, i.e.: $\sigma^v < \sigma^a$, $\tau^v > \tau^a$; to mimic the different spatial and temporal acuity. All other parameters are set at the same value in the two

Author Manuscript

This article is protected by copyright. All rights reserved

unisensory layers (see Table S1 for parameter values and the Supplementary Material for parameter assignment criteria).

Simulations and Behavioral Indexes

As stated above, we designed this model to help identify some neural mechanisms involved in multisensory perception such as the spatial detection of the external stimuli and the discrimination of the number of external sources. To this aim, we simulated different behavioral experiments of sensory detection tasks and compared the results of the model with data present in the literature (Bertelson & Radeau, 1981; Wallace *et al.*, 2004; Wozny *et al.*, 2010; Odegaard *et al.*, 2015; Rohe & Noppeney, 2015b).

In the following, we sum up briefly how data were obtained in these different behavioral experiments, to allow the reader to appreciate the similarities and differences with the model.

(Odegaard *et al.*, 2015) and (Wozny *et al.*, 2010) adopted the same experimental set-up. Participants provided the perceived locations of visual and auditory stimuli, presented alone or combined in different (congruent or incongruent) positions along the horizontal axis. The evaluation of a common cause was based on the spatial disparity between the visual and the auditory percepts: distances less than 1° were considered as an indication of a common cause, distances greater than 5° suggested independent sources.

(Rohe & Noppeney, 2015b) investigated the effect of the stimulus reliability on the sensory perception and causal inference problem. Visual stimuli were clouds of 20 dots with a vertical SD of 5.4° and the horizontal SD set to four levels: 0.1°, 5.4°, 10.8°, or 16.2°. Participants were presented with synchronous, spatially congruent or discrepant visual and auditory signals, and performed two tasks: spatial localization of the auditory signal and common source judgment.

(Wallace *et al.*, 2004) analyzed spatial localization, causal inference problem, and temporal window of integration. A sound was presented either alone or combined with a visual input at different spatial and temporal disparities. The participants indicated the perceived auditory location and if the stimuli were produced by a common source (report of unity).

(Bertelson & Radeau, 1981) analyzed perceptual bias and causal inference problem. A ventriloquist paradigm was used where both the visual and auditory stimuli were presented alone or combined in different spatial configurations. The participants evaluated the spatial location either of the visual or the auditory stimulus and if they had the same or different origin.

Author Manuscript

This article is protected by copyright. All rights reserved

In the following, we describe the different set of simulations performed to test the model. Results for every input configuration were evaluated as the mean response over 1000 repetitions of the same task.

In a first set of simulations, we compared the spatial accuracy of the model in unisensory (auditory alone and visual alone conditions) versus multisensory conditions (see below).

Subsequently, we performed simulations in multisensory conditions, with auditory-visual stimuli presented to the network, with different spatial displacements. To obtain these configurations, we always kept the visual stimulus fixed in a specific position of the space and shifted the position of the auditory stimulus, starting from a configuration where the two stimuli are presented in the same position (i.e. stimuli coincident in space) to the case of the auditory input 20° far apart (i.e. stimuli spatially segregated).

For each stimulus configuration, the behavior of the network has been analyzed in terms of:

i) The "*Report of Unity*", referring to how often the network identifies a common cause (C = 1, i.e. one peak above threshold in the multisensory area), or two different causes (C = 2, i.e. two peaks above threshold in the multisensory area) for the two stimuli, as described by Wallace and colleagues (Wallace *et al.*, 2004). This index has been plotted versus the spatial disparity of the multisensory inputs, to identify the likelihood that the model integrates or segregates two multisensory stimuli at different distances.

ii) The "*Auditory Perception Bias*", referring to the bias in the perceived position of an auditory stimulus when presented along with a visual stimulus in a different portion of the space. This index has been computed as the spatial disparity between the real position of the external auditory stimulus, and the position evaluated by the model (i.e. the barycenter of the evoked activity in the auditory area), divided by the distance between the real auditory and the real visual stimulus. First, it has been computed in general conditions, i.e. without taking into account the number of sources identified by the network. Then, the computation has been evaluated separately in the two cases of a common cause (C=1), and different causes (C=2), in order to investigate the relationship between the perceived auditory localization and the number of inferred causes. Because the visual position is only barely affected by sounds, due to the higher visual acuity, the visual perception bias has not been reported. Results are then compared with those reported in (Bertelson & Radeau, 1981; Wallace *et al.*, 2004; Rohe & Noppeney, 2015b).

iii) The "*spatial distribution of the auditory position*", i.e., how the auditory position estimate varies around its mean value. In particular, we compared the spatial distribution of the auditory position in the unisensory versus the multisensory case, with coinciding auditory and visual stimuli. Then, we analyzed the distribution of the perceived position of the auditory stimulus at each cross-

Author Manuscript

This article is protected by copyright. All rights reserved

modal configuration to assess how the distance between the stimuli affects the spatial evaluation of the auditory input. We compared these results with those reported by (Wallace *et al.*, 2004) and by (Odegaard *et al.*, 2015).

Subsequently, in order to unmask the role of the main mechanisms in the model, we performed a sensitivity analysis. To this end, we repeated the same set of simulations described above, but, in a first set of simulations, we varied the effectiveness of the synaptic mechanisms in the model (cross-modal synapses, lateral synapses within the unisensory and multisensory area, and feedforward synapses). The aim of this set of simulations is to stress the influence of these mechanisms on the causal inference process and to identify how a change in their parameters may affect the results.

Second, we modified the standard deviations of the auditory and the visual stimuli and the added noise presented to the network, in order to understand how manipulating the spatial reliability and the uncertainty of sensory stimuli could affect the perceptual abilities of the network and the solution of the causal inference problem. For these simulations, we only present results about the report of spatial unity and the auditory bias.

Finally, we realized some additional set of simulations to better characterize the abilities of the model. First, we compared the model behavior in case of modality specific and cross-modal stimuli. Second, we evaluated the ability to solve the causal inference problem and the integrative capabilities of the network in the temporal domain.

The interested reader can replicate the simulations described above by using the Matlab files linked to this paper and uploaded in the EJN repository. Specifically, results presented in Figures 4 , 11, 12, 13 were obtained by using the CI_model.m file; Figures 2, 3,-5,-6,-7,8, 9, 10 were obtained from the CI_model_macro.m file.

Results

Simulations with no-spatial disparity

Figure 2 displays a comparison between the spatial distribution of the auditory estimate in the unisensory case versus the auditory-visual case (when the auditory and visual stimuli are coincident). The results are compared with behavioral data. In both cases (unisensory and multisensory) the auditory estimate exhibits a negligible bias (a mean value close to zero). The distribution of the estimate becomes much more precise in the multisensory condition compared with the unisensory condition. The patterns and the SDs are in good agreement with the behavioral findings.

Author Manuscript

This article is protected by copyright. All rights reserved

A summary of the SD in the unisensory auditory, unisensory visual, and cross-modal conditions (without AV disparity) is presented in Table I. In unisensory conditions, the SD of the auditory estimate is significantly greater than the SD of the visual estimate, reflecting poor spatial accuracy. In cross-modal (i.e. multisensory) conditions, we can observe a significant improvement in the SD of the auditory localization error (which falls from approximately 6.4° to 2.5°) and a further small improvement in the visual localization accuracy (from 1.03° to 0.94°). For what concerns localization in the second layer, this is always a little worse than localization in the visual unisensory layer and remains practically the same in unisensory visual and multisensory conditions. Indeed, the role of the causal inference layer in the model consists especially in the evaluation of the number of estimated sources, rather than in evaluating the position of these sources. The true benefits in the spatial localization occurs in the unisensory layers, where the position estimates follows the rules of a Bayesian estimate (but see also (Ursino *et al.*, 2017)) and not in the second layer. This is a model shortcoming, since many studies demonstrate that the integrated percept is more reliable than the unimodal percepts, according to the principles of Bayesian estimate.

Figure 3 displays the spatial distribution of the auditory perception, in cases of AV stimuli, computed separately for cases C=1 (a single perceived cause) and C=2 (two perceived sources). The results are then compared with the behavioral findings of (Odegaard *et al.*, 2015) and (Wallace *et al.*, 2004).

It is worth noting that the spatial distribution of the auditory percept in the model is comparable to the behavioral data from Wallace et al. (2004): wider when C=2, and much more restricted in case of a common source inference (C=1). The model shows some differences with Odegaard et al. (2015) for the common cause case. This could be ascribed to the criteria utilized to evaluate the number of perceived input sources in the latter work, where a perceived distance less than 1° between the auditory and visual stimuli signaled a common cause evaluation, while stimuli perceived at a distance greater than 5° signaled independent sources.

Simulation with A-V spatial disparity

To explain more accurately how the model processes a multisensory configuration and deals with the causal inference problem, Figure 4 displays the results from two exemplary simulations. In both cases, the network receives a visual input in position 0° and an auditory stimulus 10° to the right of the visual input. However, due to the noisy nature of the sensory perception, the results are very different in the two cases.

Author Manuscript

This article is protected by copyright. All rights reserved

In the first simulation (panel A), the activities in the two unisensory regions are largely superimposed (i.e., they cover proximal space positions). Consequently, the multisensory region receives an excitation targeting the same neurons and, as shown in the upper panels, presents only one peak of activity above threshold. Therefore, the model 1) infers from the multisensory activity that both stimuli are produced by the same external event (C=1); and 2) infers from the activities in the unisensory areas that the auditory and visual inputs are coming from close portions of the external world. This also produces the strong positive bias (see figures 5 and 6 below) in the case of an AV configuration with a relative distance of 10°, wherein the model infers a common cause for the two stimuli.

In the second simulation (Panel B), even if the real distance between the stimuli is the same, the model identifies it to come from two different sources (C=2 in the multisensory area). In fact, in this case, the auditory and visual activities emerge mainly in different portions of the unisensory areas due to noise, and the evoked activity in the multisensory region presents two distinct peaks above threshold. Moreover, it is worth noting that the auditory percept (i.e. the barycenter of the activity in the auditory region) is more distant from the visual percept than the real input distance. Thus, in this case, we can say that the cross-modal effect of the visual stimulus produced a negative bias on the perception of the auditory position (Figure 6 when C=2).

An additional result emerges from these simulations. In cases of two stimuli eliciting superimposed activities in the unisensory regions (panel A), the network infers a common cause very quickly: after just 10 ms, the multisensory region presents a single peak above threshold. Conversely, in cases of stimuli evaluated as produced by separate sources (panel B), the time interval necessary for having two distinct peaks above threshold in the multisensory area is longer (about 20 ms). This prediction could be tested in future behavioral experiments. Of course, the values reported above (10 ms and 20 ms, respectively) crucially depend on the time constant used for the neural units in this model. Hence, they are only representative of a trend and can be modified (for instance, increased) using a different time constant value.

Figure 5 shows the report of unity and the total auditory perception bias versus the auditory-visual distance, in cases of spatially separate multisensory inputs.

The Report of Unity (Panel A, Figure 5) shows that, for an audiovisual disparity smaller than 8°, the model identifies the two stimuli as produced by the same cause in more than the 80% of simulations. In the interval tested, this frequency decreases linearly with the distance: the greater the distance between the auditory and visual stimuli, the more likely the model identifies two separate sources. These model results are in line with human performances.

Author Manuscript

This article is protected by copyright. All rights reserved

The model auditory bias is quite constant (about 70% of the AV discrepancy) in cases of audio-visual disparity smaller than 10°. The bias decreases linearly as the discrepancy increases. A similar pattern has been shown by Bertelson and Radeau (Bertelson & Radeau, 1981), Wallace and colleagues (Wallace *et al.*, 2004) and Rohe and Noppeney (Rohe & Noppeney, 2015b). The last set of authors, however, used different experimental conditions for the auditory localization tasks: Wallace and colleagues presented multisensory stimuli with different spatial configurations and different temporal disparities. Rohe and Noppeney modified the spatial configurations of the stimuli and the visual reliability. In Figure 5, we compare the model's results with the experimental conditions that are more similar to our configurations.

Figure 6 shows the Auditory Perception Bias evaluated separately for the cases of common cause inference (C=1, solid line), and cases where the model identified separate sources for the stimuli (C=2, dashed line). It's worth noting that when the model infers a common cause for both stimuli, the perceived position of the auditory stimulus is greatly affected by the presence of the visual input, resulting in a bias towards the visual position that is greater than 75% of their real distance, even for stimuli presented at distances greater than 16°. Conversely, when C = 2, the perceived position of the auditory stimulus presents a negative bias at small AV distances, i.e., the model perceives the auditory position more distant than in the reality. Comparing these results with experimental data (Bertelson & Radeau, 1981; Wallace *et al.*, 2004; Rohe & Noppeney, 2015b), it is worth noting that, in cases of common cause inference (C=1), the model presents 1) an auditory bias almost constant with the AV distance (a behavior in line with human data), and 2) a smaller bias than reported in Wallace et al. (Wallace *et al.*, 2004) (highest value = 80% in simulations versus 93%), but larger than Rohe and Noppeney (Rohe & Noppeney, 2015b) and Bertelson and Radeau (Bertelson & Radeau, 1981). Conversely, when the model identifies different sources (C=2), the amount of the bias obtained from the simulations is comparable with human performances reported by Wallace and colleagues for small spatial disparity, and by Bertelson and Radeau at distances greater than 8°, but it shows an opposite behavior with respect to Rohe and Noppeney. These differences will be analyzed and discussed in the following (see the sensitivity analysis).

Figure 7 displays the distribution of the perceived auditory position as a function of AV spatial disparity. The auditory percept distribution is compared with the positions of the external auditory (vertical dashed lines) and visual (vertical solid lines) inputs.

It is worth noting that for stimuli at a close distance (AV distance ≤ 10°), the model generates an auditory distribution mainly centered at the position of the visual stimulus: the auditory spatial perception is greatly affected by the more accurate visual input. Conversely, when the AV distance

Author Manuscript

This article is protected by copyright. All rights reserved

is in the range 10°-20°, the model presents a spatial distribution of the perceived auditory position characterized by two peaks: the first is close to the real position of the visual stimulus, while the second is centered on the real position of the auditory input.

Figure 7 is important since, as shown in (Wozny *et al.*, 2010), the distribution pattern of the estimated auditory position provides a characterization of the decision-making strategy for each individual, i.e., allows discrimination between different Bayesian models of causal inference. In particular, the bimodal distribution of Fig. 7 has been observed by Wozny et al. (Wozny *et al.*, 2010) in about 82% of subjects and is in accordance with behavioral data obtained by Odegaard and colleagues (Odegaard *et al.*, 2015). In order to better understand the spatial distribution of the auditory perception, we computed the SD of this distribution separately in the cases C = 1 and C = 2, and plotted these values versus the AV spatial disparity (Figure 8). The results were then compared with those by Wallace et al. (Wallace *et al.*, 2004). The figure shows that the auditory standard deviation is always greater in cases of two perceived sources when compared with the single source estimation. Moreover, when C = 2, the SD is maximal when the two cross modal stimuli are coincident (depicted in Figure 3) and decreases with larger AV discrepancies. Conversely, the SD is quite small when a single source is perceived because, in this case, the more precise visual stimulus drives the perception. It increases only with large AV discrepancies. These patterns agree with those by Wallace et al. (Wallace *et al.*, 2004) fairly well, although the SDs of these authors are somewhat greater.

Sensitivity analysis

Finally, we performed a sensitivity analysis on the main mechanisms incorporated in the model to reveal how they affect the report of unity and, when applicable, the auditory localization bias. The analysis is subdivided into three parts. First, we consider the synaptic mechanisms operating in the unisensory areas (Fig. 9 and Fig S1 in the Supplementary Material). Then, we analyze the role of the accuracy and noisy component of the input stimuli (Fig. 10 and Fig. S2). Finally, we analyze the synaptic mechanisms working in the multisensory layer (see fig, S3 and S4 in the Supplementary Material'.

Results show that the most influential parameters are those in the unisensory layers, while parameters in the multisensory net are less influential. Hence, the latter results are reported only in the Supplementary Material.

Fig. 9a describes the effect of a variation in the strength of the cross-modal synapses, on the Report of Unity and on the auditory localization bias. The results suggest that the direct connectivity between the two unisensory regions plays a pivotal role in producing a single cause inference. If this connectivity is too weak (0.7 instead of the basal value as high as 1.4), the

This article is protected by copyright. All rights reserved

Author Manuscript

probability that the model infers a single cause (C = 1) is drastically reduced. Conversely, increasing the efficacy of these synapses causes a strong increase in the single source inference, with an almost 100% report of unity at small AV distances and about a 70% report of unit even at 20° AV disparity. Similarly, the auditory bias is dramatically reduced, if these synapses are weakened.

Fig. 9b analyzes the role of the strength in the lateral competitive mechanisms in the unisensory areas. The results show that stronger competition in the unisensory areas reduces the Report of Unity and the auditory localization bias. Conversely, weaker lateral synapses are associated with a greater Report of Unity and with a greater localization bias, i.e., the subject exhibits a greater tendency to unify the auditory and visual signals into a single percept (see also Fig. S1 for additional results).

In general, the mechanisms working on the unisensory areas are much more influential on the causal inference than those entering the multisensory layer. Their essential role can be explained as follows: 1) The cross-modal connections are the only mechanism in the network implementing a reciprocal influence between the two stimuli localizations. Therefore, this mechanism mostly controls the causal inference. Strong cross-modal connections increase the probability that the two localizations get close to each other, increasing the report of unity and auditory bias. 2) The lateral competition within the unisensory areas modulates the attraction effect mediated by the cross-modal synapses. Stronger lateral competition (implemented by increasing the strength of the lateral connections), reduces the activation bubble in each layer, decreasing the influence of cross modal synapses and so the report of unity and the auditory bias.

Finally, Fig. 10 shows results of a sensitivity analysis on the noisy component and the accuracy of the sensory stimuli. It is interesting to note that by modifying the level of the noise added to the network, the model reproduces the different behavioral data from Wallace and colleagues (Wallace *et al.*, 2004) and Rohe and Noppeney (Rohe & Noppeney, 2015b) regarding the auditory perceptual bias in case of independent sources. The former found a strong negative bias in this condition (C=2) for stimuli presented at distances less than20°. Conversely, the latter found a positive bias in case C=2, under the same spatial configurations of the stimuli. As illustrated in Figures 10a and 10b, our model is able to reproduce Wallace's data with a low level of noise, while a high noisy condition is necessary to obtain the positive bias showed by Rohe and Noppeney in case of multisensory stimuli at small distances.

Unisensory vs. multisensory casual inference

This article is protected by copyright. All rights reserved

Author Manuscript

It is worth-noting that the present model can perform causal inference not only in multisensory conditions, but also in case of unisensory inputs. In order to clarify this important aspect, we performed some simulations comparing casual inference in unisensory vs. cross-modal conditions (Fig. 11).

Fig. 11 shows that two unisensory visual inputs at a 20° distance are perceived as produced by two independent causes. Due to the lateral competition in the unisensory and multisensory regions, this inputs configuration produces weak responses in the multisensory layer, just above the detection threshold. Conversely, two auditory inputs at the same distance are perceived as originating from a single cause, located in between the original positions of the two stimuli. This is a consequence of the poor auditory spatial resolution. Finally, two multisensory stimuli at 20° distance elicited strong activities in the areas of the model, showing a strong multisensory enhancement, and the stimuli are perceived as coming from a single cause, but located close to the visual position.

To stress the significant role played by multisensory integration in the model, we tested the network with unisensory and multisensory stimuli with different levels of intensity. Results, reported in Figure 12, highlight the benefits that multisensory integration can exert in case of weak but congruent stimuli, whereas this benefit is quite irrelevant in case of strong inputs. The figure shows the different effect exerted, on the second layer, by: a) a weak auditory stimulus alone, b) a strong auditory input alone, and c) two weak congruent multisensory stimuli. In the first case, the network fails to infer any cause, since the activity evoked in the causal-inference layer by a weak auditory input is sub-threshold. However, if the weak auditory stimulus is paired with a weak visual input, activities in the unisensory layers are reinforced, and the second layer presents a peak of activity above threshold, inferring a single cause.

Results in Figs. 11 and 12 emphasize the perceptual advantages of multisensory cues and clarify that the model can perform causal inference also in case of two unisensory inputs.

.

Temporal aspects

Finally, the last simulations investigate the temporal aspects of multisensory integration and causal inference problem. To this end, we performed two simulations with multisensory stimuli, placed 20° apart, varying the stimulus onset asynchrony (SOA, i.e., the temporal lag between the stimuli). The stimuli lasted 100 ms in each condition. Results are reported in Figure 13. In the first case, with a SOA = 75ms (Fig. 13a), the causal inference layer presents only a single peak of activity in response to the multisensory stimuli, lasting above the detection threshold for the entire

Author Manuscript

This article is protected by copyright. All rights reserved

duration of the stimulation. This signals that the network infers a common cause for the perceived stimuli. Moreover, it is interesting to note that, even after the removal of the visual input (105ms), we can observe a strong positive bias of the auditory percept, attracted by the visual stimulus, and a significant integration (i.e., multisensory response exhibits a clear enhancement) in the multisensory layer. Conversely, if the SOA is increased at 100 ms, the network identifies two distinct peaks above the threshold in different instants (in the temporal domain). Moreover, we do not observe an auditory bias in the unisensory area, which means that the auditory percept is not affected by the previous visual input, and the multisensory layer does not present enhancement.

It is worth noting here, that the temporal window for integration and the solution of the causal inference problem is strongly related with the duration of the stimuli and the time constant of the differential equations, and can be modified by changing these parameters.

Discussion

Comparison with previous models - Several studies in recent years have focused on the "causal inference problem", i.e., the problem of deciding whether two stimuli are produced by the same source or by distinct sources. For what concerns multisensory integration, the theoretical aspects have been assessed in several pivotal papers (Pouget *et al.*, 2003; Ma *et al.*, 2006; Körding *et al.*, 2007; Shams & Beierholm, 2010; Wozny *et al.*, 2010; Ma & Rahamati, 2013; Pouget *et al.*, 2013), under the assumption that the brain realizes a near optimal Bayesian estimate. Behavioral data confirm these predictions, showing that the brain behaves quite optimally in a variety of multisensory tasks (Shams *et al.*, 2005; Körding *et al.*, 2007; Wozny *et al.*, 2008; Wozny & Shams, 2011; Samad *et al.*, 2015).

However, despite these important recent contributions, the knowledge of the neural mechanisms able to produce a Bayesian estimate is still poor. It has been proposed that the brain exploits a "neural population code", i.e., that the estimate is extracted from the activity of a population of neurons, which code for the property under examination (in the present exempla, the position of the stimulus) and implement the probability functions required. However, the biological neural network that can realize this kind of inference and the synaptic organization that is more appropriate are still unresolved determinations. In particular, we are aware of only a few biologically inspired neural networks that model the audio-visual causal inference process (Ma & Rahamati, 2013; Yamashita *et al.*, 2013).

Author Manuscript

This article is protected by copyright. All rights reserved

Ma and Rahmati (Ma & Rahamati, 2013) analyzed a possible neural implementation for causal inference, using probabilistic coding, by translating the Bayesian decision rule directly into a neural network. However, they reached the conclusion that the resulting architecture is quite unrealistic.

Yamashita et al. (Yamashita *et al.*, 2013) built a recurrent network for multisensory integration, and found that the network can infer the causal structure and reproduce the localization bias of the perceived position. Their model has some elements in common with the present one. For instance, the role of cross modal synapses in our model, reflecting prior knowledge, is played by recurrent synapses in the model by Yamashita. Also similar to our model, the number of causes is distinguished in Yamashita's model on the basis of the number of peaks in the multisensory layer, reflecting a lateral competition. However, important differences exist. In the model by Yamashita et al., all computations are performed within a single multisensory layer. As a consequence, in case of a common cause inference (C = 1), the auditory and visual perceived positions are always identical. In contrast, in our model (as in the brain (Rohe & Noppeney, 2015a)) the positions are computed within the upstream unisensory layers. Therefore, even in cases of a single source estimation, the auditory localization is not always superimposed on the visual localization, although they are often close; this result appears to agree with behavioral data. In particular, in Bertelson and Radeau (Bertelson & Radeau, 1981), when a subject reports a single cause, the auditory shift is about 3.66° at a 7° separation distance, and 8.15° at a 15° separation distance. Given that the shift in the visual localization is typically small (e.g., 0.55° and 1.05°, for these disparities, respectively), the auditory and visual perceived positions do not appear to coincide. A similar conclusion can be reached looking at the results by Rohe and Noppeney (Rohe & Noppeney, 2015b). The authors report that, when a common cause is inferred, the auditory localization bias is about 60% or 80% of the audio-visual distance, using a visual stimulus with physiological reliability. Once again, given that the visual localization bias is generally small (usually smaller than 20%), we can conclude that the auditory and visual perceptions do not coincide. These results can be simulated quite well with our model, taking also parameter variability into account (see Fig. 9), but cannot be simulated by Yamashita's model, which postulates the superimposition of the auditory and visual localizations when C = 1. Furthermore, our hierarchical organization is closer to recent findings (Rohe & Noppeney, 2015a) and reflects the well-documented existence of cross-modal links between unisensory areas (Ghazanfar & Schroeder, 2006; Musacchia & Schroeder, 2009; Recanzone, 2009). Moreover, the model by Yamashita et al. includes a weak divisive normalization to obtain the results (an operation not necessary in the present model).

Other recent models analyzed the way multisensory integration can be realized. Parise and Ernst (Parise & Ernst, 2016) proposed a model based on single computational units which work as a

Author Manuscript

This article is protected by copyright. All rights reserved

multisensory correlation detector (MCD). Each unit filters the individual sensory signals and then combines them linearly to detect correlation and time lag, thus performing synchrony and temporal order judgments. Optimal cue integration in space is then achieved via a population of MCDs, each receiving inputs from a limited spatial region. Compared with our model, Parise and Ernst realize integration completely in a multisensory layer (since units are intrinsically multisensory) and includes a time lag detector explicitly. Conversely, in our model multisensory integration is performed at two steps, while the time lag is not explicitly detected (but simply affects network output via the superimposition of the inputs, see Fig. 13). Moreover, our model incorporates lateral inhibitory mechanisms and non-linear saturation, which are important in multisensory integration (see (Ursino *et al.*, 2014)).

A different approach is used in Zhang et al. (Zhang *et al.*, 2016). These authors developed a model in which multisensory integration is performed by many interconnected multisensory areas, while connectivity among these areas reflect prior knowledge about similarity. This approach is very promising; however, multiple multisensory areas are probably required to process multi-feature object representation, not merely position [in this regard, the model by Zhang et al. resembles our recent model of semantic memory (Ursino *et al.*, 2015), in which each area codes for a different feature]. For what concerns position, the two unisensory areas in our model (which become multisensory because of reciprocal cross-modal synapses) resemble the interconnected multisensory areas by Zhang et al. Indeed, in a recent theoretical study (Ursino *et al.*, 2017) we demonstrated that the cross-modal synapses between these unisensory areas can be trained to reflect an a priori knowledge on the co-occurrence of auditory and visual stimuli, i.e. prior spatial similarity.

Model simulations vs. real data - Despite its parsimonious structure, the present model can reproduce several characteristics of the AV causal inference, such as the dependence of the number of estimated sources on the AV distance, the ventriloquism effect, the standard deviation of the acoustic localization and its bimodal distribution. These results emerge from the interaction among three basic neural mechanisms. Each of them is biologically plausible.

i) Two unisensory areas (auditory and visual) code for the position of the two stimuli and reciprocally exchange their information via spatially confined cross-modal synapses. This is the fundamental mechanism in the model, able to account for the ventriloquism effect. In a previous paper, we further demonstrated that the same mechanism can account for other multisensory illusions as well, such as illusions in the temporal domain, where auditory beeps affect the perception of the number and duration of visual flashes (Cuppini *et al.*, 2014). Moreover, we

This article is protected by copyright. All rights reserved

Author Manuscript

recently demonstrated that these cross-modal synapses can be learned from experience in a multisensory environment, reflecting the presence of multisensory stimuli in temporal and spatial proximity (Ursino *et al.*, 2017).

Recent studies have shown that cortical areas, usually associated with modality-specific sensory processing, are stimulated from other senses (see Ursino et al., 2014, for a Review). Bizley, King et al., in a series of works (Bizley *et al.*, 2007; Bizley & King, 2008; Bizley & King, 2009) have shown that neurons in the ferret auditory cortex receive influence from visual stimuli, and that multisensory integration is common to all auditory cortical areas, with a prevalence in higher ones (Bizley *et al.*, 2007; Bizley & King, 2009). In some neurons, pairing visual and auditory inputs can increase the available spatial information (Bizley & King, 2008). Moreover, in agreement with the present model, the visual inputs to auditory neurons exert subthreshold influences, able to modulate the responses to sound (Bizley & King, 2009). Although this kind of influence might also derive from a feedback from multisensory areas back to unisensory regions, neural tracer injections revealed direct input from visual cortex into the auditory cortex (Bizley *et al.*, 2007), as assumed in our model.

Other works reveal influences from the auditory into the visual cortex. Data by Iurilli et al. (Iurilli *et al.*, 2012) suggest the existence of inhibition from auditory to primary visual neurons in the mouse, via cortico-cortical connections; the authors suggest that this auditory influence may reduce potentially distracting sensory processing in the visual cortex. This mechanism may be a consequence of cross-modal excitation between the two areas (as in our model) joined with lateral inhibition within unisensory areas. Ibrahim et al. (Ibrahim *et al.*, 2016) analyzed cross-modal modulation from sounds to orientation selective visual neurons in the mouse primary visual cortex; the effect was a decrease in the average response, but an increased response at the cell preferred orientation. Although we did not account for orientation selectivity in our model, these data are consistent with the existence of an excitatory cross-modal auditory-visual link sharpened by lateral inhibition.

ii) A second mechanism consists of the convergence of auditory and visual activity towards a downstream area where integration is performed. We assume that the main role of this layer is to discriminate between two sources or a single source for the observed activities. To this end, activity in this area is compared with a detection threshold, to decide whether, in a given portion of the space, there is enough activity to infer the presence of a reliable source of information or not. It is worth noting that the second layer, although multisensory, can also be used to infer the presence of one or two causes also in within modal conditions (see Fig. 11). However, in these conditions the causal inference problem may be solved directly in the unisensory layers.

Author Manuscript

This article is protected by copyright. All rights reserved

iii) The two previous mechanisms could not work properly without the presence of a third mechanism, i.e., competition within the same layer, realized with a Mexican Hat disposition of synapses. In particular, we used a large inhibition in the unisensory areas, assuming that, here, attention is focused on a single stimulus, but a shorter inhibition in the multisensory area, to favor the emergence of one or two simultaneous activities, i.e., one or two sources of information.

A further assumption in the model, not implemented via internal mechanisms, but with the external inputs, is that the visual stimulus is spatially much more precise than the auditory one, but with a lower temporal acuity. Apart from these differences, the network in Fig. 1 is completely symmetrical. We chose to use a perfectly symmetrical network to keep the number of ad hoc assumptions to a minimum. It is probable that non-symmetrical synapses exist in real biological networks, but, here, we demonstrated that this is not essential, and differences in audio-visual processing derive merely from differences in spatial reliability of the stimuli.

Considering the main mechanisms delineated above, source estimation in the model is performed via the following computational steps: 1) the presence of lateral inhibition within the unisensory areas leads to the formation of two activation bubbles (one in each area) close to the true position of the stimuli. The presence of noise, however, may modify the position, especially for what concerns the auditory one, characterized by poor spatial resolution. 2) When the two bubbles are close enough (due to the small distance between the stimuli and/or the effect of noise), they are attracted reciprocally by the action of cross-modal synapses. Moreover, due to the poor spatial localization of the auditory stimulus, the visual input exerts a stronger attraction on a large portion of the auditory network (where many neurons, although silent, have an excitation close to the activation threshold). Conversely, since only a small portion of the visual area is close to the activation threshold, the effect of the auditory input on the visual area is generally modest. Here we did not report the visual shift in the interest of space, but it is typically quite negligible (less than 1°, see also (Magosso *et al.*, 2012; Rohe & Noppeney, 2015b)). 3) Finally, when the two activities in the unisensory areas are quite superimposed (because of the initial AV distance, the effect of noise and the cross-modal attraction), a single activation peak is formed in the downstream region. In this case, the activities of neurons that receive congruent inputs are strongly enhanced. Conversely, if the two activities exhibit only a modest overlap, the effect of lateral inhibition in the downstream area prevails, leading to the formation of two distinct peaks, surrounded by an inhibited zone. This may even result in a repulsive effect, with the two peaks farther than the original distance.

We showed that these simple mechanisms could explain most behavioral results fairly well. First, in case of coincident multisensory stimuli (Figure 2 and Table I), the auditory and visual localization errors decrease compared with that of the unisensory case, even if the reliability of the

Author Manuscript

This article is protected by copyright. All rights reserved

integrated percept in the second layer (Table I, third column) is a little poorer than the reliability of the visual unisensory percept. This is an important model limitation. This prediction disagrees with some studies (for instance Alais and Burr, 2004), which showed that the integrated percept is more reliable than the unimodal estimates, in agreement with the Bayes rules.

Second, the auditory localization bias, i.e., the ventriloquism effect, Figure 6, is strong when C = 1, but is quite negligible (or even negative at small AV distances) in cases of C = 2.

Third, the standard deviation of the auditory localization is always greater when C = 2 than in the case of a single source estimation, at all AV distances (Figure 8). The reason is that, when C =1 the auditory stimulus is attracted by the more precise visual stimulus, which constraints its position. Conversely, when *C* = 2, the auditory position is just moderately affected by the visual one, showing the overall spontaneous auditory localization variability.

Fourth, results of the sensitivity analysis underline that inter-subject variability, reported in the literature, may be ascribed to differences in the strength of synapses or in the noise, which, in turn, may be a consequence of prior multisensory experience (see also (Ursino *et al.*, 2017)) or of the variations in experimental set-up.

The sensitivity analysis also suggests that the mechanisms working on the multisensory layer (the feedforward synapses and the lateral competition) have a lesser impact on the Report of Unity than the mechanisms acting on the unisensory layers. Accordingly, an important conclusion of our model is that the localization bias, computed in unisensory areas, is a pre-condition for the causal inference, and not viceversa. First, the unisensory layers compute the attraction between the two signals based on their accuracy, i.e., the standard deviation of the inputs, the lateral competition and the presence of cross-modal synapses (reflecting previous knowledge). Only subsequently, based on this shift, the second layer infers the presence of one or two causes.

Although we are not certain that this is the best strategy (compared with a strategy that first tries to solve the causal inference problem, and only later produces a bias) we claim that it finds some support in the general idea of auto-association, exploited in many neural-network models. In our model, cross-modal synapses implement an auto-association network which tries to recover lacking information on the basis of past experience (in this case, past experience is the spatial proximity of audio and visual stimuli when C = 1). The same mechanism (i.e., auto-association via recurrent synapses) can be exploited in more general situations, whenever the brain needs to merge different pieces of information into a single percept. For instance, in a recent model of semantic memory (Ursino et al., Neural Networks, 2015) we linked features representing a single object via auto-associative synapses, thus allowing object recognition even in presence of a partial cue. In conclusion, we think that auto-association via recurrent synapses is a powerful way to favor

Author Manuscript

This article is protected by copyright. All rights reserved

solution of the causal inference problem, restoring the information which is expected to occur in the C = 1 case. If this information cannot be restored, the C = 2 case can be assumed as more reliable.

Model predictions. Various testable predictions derive from our simulations. First, based on the above statement, the response time for a spatial localization test (when an observer decides the position of the auditory stimulus) should be faster than the response time of a causal inference test, i.e., when the observer decides on one or two causes. In other terms, the first decision anticipates the second, and not viceversa.

Similarly, as illustrated in Figure 4, the model predicts that the causal inference response is faster when C = 1 than in cases of inferring two independent sources. In fact, in the first case, the superimposition of two congruent activities in the multisensory area allows a quicker attainment of the detection threshold. This may represent a testable prediction, validated in future experiments. Indeed, this result substantially agrees with the present knowledge of multisensory integration. Data by Arnal et al. (Arnal *et al.*, 2009) suggest the presence of temporal audio-visual facilitation in temporal processing.

Furthermore, the model predicts that the distribution of the auditory localization exhibits a bimodal pattern, which reflects the C = 1 and C = 2 inferences. Moreover, the two modes are not completely distinct, but exhibit a certain superimposition even at large AV distances (Figure 7). This distribution agrees with behavioral data observed by (Wozny *et al.*, 2010) in more than 75% of cases.

Model limitations and future lines - Of course, the present model is still preliminary; we followed a parsimony principle to lessen its complexity and reduce the number of mechanisms involved in order to focus on general ideas rather than on a complete accurate description of all phenomena. Several limitations can become the target of future improvements.

First, we assumed that the sensory reliability of the stimuli is independent of the position. Conversely, it is well known that the perception of space for auditory and visual neurons strongly depends on the azimuthal coordinate. This phenomenon has been extensively investigated recently by Odegaard et al. (Odegaard *et al.*, 2015). We believe that inclusion of this dependence in the model (so that peripheral neurons are less accurate than the central ones) can affect the results, especially at large AV distances.

Second, noise in real biological networks can originate at different levels, not only in the input stimuli.

Author Manuscript

This article is protected by copyright. All rights reserved

Third, these results demonstrated that the mechanisms implemented in the unisensory layers exert a major role to generate sensory perception and solve the causal inference problem; but we are aware that in real biological networks feedback projections are also present from higher order multisensory brain regions to primary cortical regions. Future experiments and simulations will be conducted to analyze a possible role of these projections in the brain sensory processing.

Fourth, we used the same parameters for all simulations, considering noise as the only cause of trial-by-trial variability. Conversely, behavioral data are obtained on different subjects, which significantly differ as to their responses. As stated by Hairston et al. (Hairston *et al.*, 2003), "performances proved to be highly variable among subjects. The source of this inter-subject variability is not immediately clear". This inter-subject variability may involve a difference in ability to locate visual targets versus auditory targets, variability in strength of the synapses (see the sensitivity analysis in Figure 9 and Figures in the Supplementary Materials) which, in turn, may reflect variability in previous multisensory experience, according to Hebbian learning paradigms (Ursino *et al.*, 2017), or variability in detection threshold. A future experiment may use the network separately on different subjects to fit individual observers' data with individual parameter estimates. This will provide a more accurate model validation and a deeper understanding of the neural origin of individual variability.

Finally, for the sake of simplicity we mimicked localization in the auditory area in the same way as visual localization. However, such topological organization is not present in the auditory cortex (although the superior colliculus has spatial auditory map; however, it is not likely to be the locus of localization). Hence, more reliable models for the auditory spatial representation should be developed in future work. In particular, some authors hypothesized that space is represented in the auditory cortex in a distributed way (Stecker & Middlebrooks, 2003) and that sound space localization is based on ensemble of cortical neurons (Middlebrooks *et al.*, 1998). These ideas are not in opposition with the present model, but their implementation will require a more sophisticated description of the auditory cortex, possibly using temporal coding with synchronized neural oscillators. In particular, an ensemble of synchronized oscillators can send their activity toward other areas in a way similar to what is performed by a single unit in our model (see (Ursino *et al.*, 2009) for an example).

Acknowledgments:

Author Manuscript

This article is protected by copyright. All rights reserved

C.C. was supported by the Italian Ministry of Education, Project FIRB 2013 (Fondo per gli Investimenti della Ricerca di Base-Futuro in Ricerca) RBFR136E24. L.S. was supported by a National Science Foundation grant (BCS-1057969) and Oculus Research.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We thank Brian Odegaard for his support in the analysis of the behavioral data from his previous work, and Angela Bubis and Thomas Botch for proof reading the manuscript.

Author contributions:

C.C. implemented the model. C.C., M.U. and L.S. analyzed the results. C.C., M.U. and E.M. wrote the manuscript.

Data Accessibility Statement

All data used for these analyses and all supporting information files can be accessed online. Data used for these analyses are in the form of Matlab files located at figshare.com

Figure Captions

Figure 1 – Structure of the network. The visual and auditory regions process external sensory stimuli. These regions are reciprocally connected through direct excitatory synapses (W^{av} and W^{va}), and send long-range feedforward projections (W^{ma} and W^{mv}) targeting the causal inference area. All these inter-area synapses are realized via Gaussian functions. The three regions in the network include also intra-area synapses, linking elements belonging to the same area. These connections are implemented by using a Mexican-hat function.

Figure 2 – Distribution of the auditory localization error simulated with the model (dark bars) and measured in the behavioral experiments (light bars, i.e., data 1, Odegaard et al., 2015) in unisensory conditions (auditory input alone: left panel) and in multisensory conditions (auditory and visual stimuli at the same position, right panel). The insert tables report the corresponding standard deviations. SDs from other two additional studies (data 2 from Rohe and Noppeney, (Rohe & Noppeney, 2015b), data 3 from Hairston et al., (Hairston *et al.*, 2003)) are also reported for comparison. In all cases, the presence of a congruent multisensory stimulus reduces the SD of the localization error, although a large variability can be observed among behavioral data.

This article is protected by copyright. All rights reserved

Author Manuscript

Figure 3 – Distribution of the auditory localization error (i.e., the difference between the perceived auditory position and the true position) computed in multisensory conditions with spatially-congruent auditory and visual stimuli. The cases of single source estimation (C = 1) and distinct source estimation (C = 2) are plotted separately. Model results (left panel) are compared with Odegaard et al. (Odegaard *et al.*, 2015) behavioral data (middle panel) and with data by Wallace et al (Wallace *et al.*, 2004) (right panel).

Figure 4 – Examples of neural activity temporal patterns. The upper panel (panel A) shows the simulation of a case where two multisensory stimuli, presenting an A-V distance as large as 10°, are considered as originating from a single cause. The bottom panel (panel B) shows an example in which the two stimuli, with the same spatial configuration, are considered as originating from separate causes (due to noise). Within each panel, the upper rows describe the activity in the multisensory layer, while the lower row the activities in the unisensory layers (dash-dotted visual, dashed auditory). In each panel, the different columns represent three snapshots of network activity at three different instant during the simulation. The left column is network behavior at the beginning of the simulation (10 ms), when the multisensory area is still scarcely active (below threshold); the middle column is an intermediate instant (20 ms), when the threshold has already been reached in the multisensory area; the right column represents the final configuration (45 ms). It is worth noting that a multisensory stimulation with similar spatial configuration can be processed in different ways by the model, as a consequence of the noisy sensory perception.

Figure 5 – Perception of unity (panel A) and auditory perception bias (panel B) as a function of AV discrepancy. Panel A reports the percentage of times the network infers a common cause during crossmodal stimulation, plotted vs. the distance between the auditory and the visual components (red line). For stimuli with a distance smaller than 8°, the network identifies a common cause in more than 80% of the simulations. For stimuli with greater distances, the percentage of identification of a common source decreases linearly with the distance. The simulations' results (red solid line) are compared with behavioral data from Rohe and Noppeney, (Rohe & Noppeney, 2015b) (green line), Wallace et al., (Wallace *et al.*, 2004) (blue line) (obtained with a similar experimental paradigm, but with an auditory-visual temporal disparity of 200ms) and by Bertelson and Radeau (Bertelson & Radeau, 1981) (magenta line). Panel B shows the Auditory Perception Bias, i.e., the difference between the perceived spatial position and the true position of the auditory input, expressed as a percentage of the audio-visual distance. The bias of the auditory perceived position is represented by the mean and the SEM, computed over 1000 simulations for each spatial input configuration.

Figure 6 – Auditory perception bias in multisensory conditions, evaluated separately when the network identifies a common cause (C=1, red solid line) or different causes (C=2, red dashed line). The results of model simulations (red lines) are compared with behavioral data present in literature (blue lines - Wallace et al., (Wallace *et al.*, 2004); green lines - Rohe and Noppeney, (Rohe & Noppeney, 2015b); magenta lines -

Author Manuscript

This article is protected by copyright. All rights reserved

Bertelson and Radeau (Bertelson & Radeau, 1981)). The network predicts that in the first case (C=1), the auditory perception is affected by a bias greater than 75% and is fairly constant across AV spatial disparities; for C=2, the auditory bias is negative for a distance smaller than 10°, i.e. the perceived auditory position is more distant from the visual input than in the reality.

Figure 7 – Distribution of the auditory localization in cross modal conditions, at different AV distances. The visual stimulus was always located at position 0 deg (continuous vertical line); the auditory position is indicated in each panel (dashed vertical line). Panels in the first row are behavioral data from Odegaard et al (Odegaard *et al.*, 2015); panels in the bottom row are from model simulations. A bimodal distribution becomes evident at large AV distances.

Figure 8 – Standard deviation of the auditory localization at different AV distances, computed by the model separately in the cases of a single source estimation (C = 1) and distinct source estimations (C = 2). For comparison data by (Wallace *et al.*, 2004) are reported, too.

Figure 9 – Sensitivity Analysis on the mechanisms operating in the unisensory areas. The figure shows the effect of changing some parameters in the unisensory areas, on the Report of Unity and the auditory localization bias, at different A-V distances. Figure 9a has been obtained with different values of the weights of the cross-modal mechanism (i.e., a change in the weight of direct synapses among unisensory areas $W^{av} = W^{va}$). Stronger cross-modal connections enhance the influence of the visual stimulus. This increases 1) the likelihood of the perception of a common source (Report of Unity) for the visual and auditory stimuli and 2) the bias of the perceived position of the auditory stimulus. Conversely, weak connectivity among the unisensory regions increases the ability of the network to identify separate stimuli also at small distances: for W0 = 0.7 the network identifies independent sources in more than 50% of the cases for stimuli placed at a distance of 10° or less. The same result is obtained only for distances greater than 16° in the basal configuration (W0 = 1.4). Figure 9b describes the effect of a change in the strength of the lateral competition mechanism, in both the auditory and visual areas [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)]. In this case, the effect is opposite with respect to the previous case: the stronger the competitive connections the lower the likelihood of the perception of a common cause for the AV stimuli and the perception bias of the auditory input. A strong inhibition among elements within the same unisensory area reduces the effect of the crossmodal input. This helps to keep segregated two stimuli placed in different spatial positions, resulting in the perception of independent input sources. In all panels, the basal condition is displayed with a continuous line.

Figure 10 - Sensitivity Analysis on the effect of noise, added to the sensory inputs, on the auditory perception bias in case of common cause (C=1, red solid line) and two independent causes evaluation (C=2, red dashed line), at different A-V distances. A) A highly noisy sensory stimulation is suitable to reproduce

This article is protected by copyright. All rights reserved

Author Manuscript

the results of Rohe and Noppeney (Rohe & Noppeney, 2015b), i.e., a positive bias in case of independent sources (C=2). Conversely, B) a low added noise in the unisensory regions is able to explain the strong negative bias identified by Wallace and colleagues (Wallace *et al.*, 2004) in case of separate cause evaluation.

Figure 11 – Unisensory vs Multisensory Integration and Causal Inference. The network is able to solve the causal inference problem for unisensory and multisensory conditions. In the figure, we presented 3 different conditions: the network was stimulated with two inputs at a distance of 20°; in case A) we used two visual inputs of the same intensity; in case B) we presented two auditory stimuli; in case C) we used a visual and an auditory input. The vertical black-dotted lines identify the original positions of the stimuli presented to the network. Blue lines are referred to the evoked activities in the visual area, green lines depict the activity in the auditory area, and red lines are used for the activity in the multisensory region. The black-dashed lines represent the detection threshold in the multisensory region. In case of visual inputs A), the network perceived the two stimuli, with no spatial perception bias (the barycenter of the evoked activities in the visual area coincides with the original position of the stimuli), as produced by independent causes (two peaks above the detection threshold in the multisensory region). In case of auditory stimulation B), the network identifies a single auditory stimulus, whose position is in between the original positions of the two stimuli (the evoked activity in the auditory region has a barycenter placed between the two original inputs, and the multisensory area shows a single peak above the threshold). In the multisensory case C), the network identifies a common cause for the two stimuli, and the perceived positions of the two inputs are very close to the original position of the visual stimulus (great auditory perception bias).

Figure 12 – The role of stimulus intensity in the multisensory integration. The figure shows the different effect, produced on the network, by: A) a weak auditory stimulus alone; B) a strong auditory stimulus alone; C) two weak congruent multisensory stimuli. In the first case, the network fails to infer any cause, since the activity in the multisensory area is sub-threshold, whereas in the second and third cases, the second layer infers a common cause. These results demonstrate that multisensory integration generates a strong benefit particularly in case of weak stimuli. This effect is present both at the level of the unisensory areas, where the evoked activities are reciprocally reinforced thanks to the cross-modal connections, and in the multisensory region, where the evoked activity presents a strong multisensory enhancement (compared with case A), which helps identifying a common cause for the two stimuli.

Figure 13 – Multisensory Integration Temporal Window. The upper panel (panel A) shows the results of multisensory stimulation with an A-V distance as large as 20°, and a stimulus onset asynchrony (SOA) as low as 75ms. The bottom panel (panel B) shows an example in which the two stimuli, with the same spatial configuration, have a larger temporal offset (SOA = 100ms). In both conditions, the duration of the stimuli

Author Manuscript

This article is protected by copyright. All rights reserved

was set at 100ms. Within each panel, the upper rows describe the activity in the multisensory layer, compared with the detection threshold (the horizontal black dashed line), while the lower row show the activities in the unisensory layers (green visual, blue auditory). In each panel, the different columns represent different snapshots of network activity at four different instants during the simulation. One snapshot is at the beginning (60 ms), when the network detected only the visual input. Two snapshots are at intermediate instants (105 ms and 120 ms) when the visual input was removed and the network received the auditory stimulus. One snapshot is close to the end of the simulations (140 ms), when only the auditory input could affect the network.

References

Alais, D. & Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, **14**, 257-262.

Andersen, T.S., Tiippana, K. & Sams, M. (2004) Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, **21**, 301-308.

Arnal, L.H., Morillon, B., Kell, C.A. & Giraud, A.-L. (2009) Dual neural routing of visual facilitation in speech processing. *J Neurosci*, **29**, 13445-13453.

Battaglia, P.W., Jacobs, R.A. & Aslin, R.N. (2003) Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A Opt Image Sci Vis*, **20**, 1391-1397.

Bertelson, P. & Radeau, M. (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys*, **29**, 578-584.

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

Author Manuscript

This article is protected by copyright. All rights reserved

Bizley, J.K. & King, A.J. (2009) Visual influences on ferret auditory cortex. *Hear Res*, **258**, 55-63.

Bizley, J.K., Nodal, F.R., Parsons, C.H. & King, A.J. (2007) Role of auditory cortex in sound localization in the midsagittal plane. *J Neurophysiol*, **98**, 1763-1774.

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G. & Ursino, M. (2014) A neurocomputational analysis of the sound-induced flash illusion. *Neuroimage*, **92**, 248-266.

Cuppini, C., Magosso, E. & Ursino, M. (Year) A neural network model of cortical auditory-visual interactions a neurocomputational analysis of the shams-illusion. In: Rosa, A.D., A.; Madani, K.;Filipe, J.;Kaprzyk, J. (ed), IJCCI 2012-4th International Joint Conference on Computational Intelligence. SciTePress-Science and Technology Publications, City. p. 639-642.

Ernst, M.O. & Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429-433.

Fetsch, C.R., Pouget, A., DeAngelis, G.C. & Angelaki, D.E. (2012) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*, **15**, 146-154.

Ghazanfar, A.A. & Schroeder, C.E. (2006) Is neocortex essentially multisensory? *Trends Cogn Sci*, **10**, 278-285.

Hairston, W.D., Wallace, M.T., Vaughan, J.W., Stein, B.E., Norris, J.L. & Schirillo, J.A. (2003) Visual localization ability influences cross-modal bias. *J Cogn Neurosci*, **15**, 20-29.

Hillis, A.E., Chang, S., Heidler-Gary, J., Newhart, M., Kleinman, J.T., Davis, C., Barker, P.B., Aldrich, E. & Ken, L. (2006) Neural correlates of modality-specific spatial extinction. *J Cogn Neurosci*, **18**, 1889-1898.

Ibrahim, L.A., Mesik, L., Ji, X.-y., Fang, Q., Li, H.-f., Li, Y.-t., Zingg, B., Zhang, L.I. & Tao, H.W. (2016) Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. *Neuron*, **89**, 1031-1045.

Author Manuscript

This article is protected by copyright. All rights reserved

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F. & Medini, P. (2012) Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, **73**, 814-828.

Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B. & Shams, L. (2007) Causal Inference in Multisensory Perception. *PloS one*, **2**, e943.

Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci*, **9**, 1432-1438.

Ma, W.J. & Rahamati, M. (2013) Towards A Neural Implementation of causal inference in cue combination. *Multisens Res*, **26**, 159-176.

Magosso, E., Cona, F. & Ursino, M. (2013) A neural network model can explain ventriloquism aftereffect and its generalization across sound frequencies. *Biomed Res Int*, **2013 (Article ID 475427)**, 1-17.

Magosso, E., Cuppini, C. & Ursino, M. (2012) A Neural Network Model of Ventriloquism Effect and Aftereffect. *PloS one*, **7**, e42503.

Middlebrooks, J.C., Xu, L., Eddins, A.C. & Green, D.M. (1998) Codes for sound-source location in nontonotopic auditory cortex. *J Neurophysiol*, **80**, 863-881.

Morgan, M.L., Deangelis, G.C. & Angelaki, D.E. (2008) Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, **59**, 662-673.

Musacchia, G. & Schroeder, C.E. (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear Res*, **258**, 72-79.

Odegaard, B., Wozny, D.R. & Shams, L. (2015) Biases in visual, auditory, and audiovisual perception of space. *Plos Comput Biol*, **11**, e1004649.

Author Manuscript

This article is protected by copyright. All rights reserved

Parise, C.V. & Ernst, M.O. (2016) Correlation detection as a general mechanism for multisensory integration. *Nat Commun*, **7**.

Pouget, A., Beck, J.M., Ma, W.J. & Latham, P.E. (2013) Probabilistic brains: knowns and unknowns. *Nat Neurosci*, **16**, 1170-1178.

Pouget, A., Dayan, P. & Zemel, R.S. (2003) Inference and computation with population codes. *Annu Rev Neurosci*, **26**, 381-410.

Recanzone, G.H. (2009) Interactions of auditory and visual stimuli in space and time. *Hear Res*, **258**, 89-99.

Rohe, T. & Noppeney, U. (2015a) Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS biology*, **13**, e1002073.

Rohe, T. & Noppeney, U. (2015b) Sensory reliability shapes perceptual inference via two mechanisms. *J Vis*, **15**, 22-22.

Samad, M., Chung, A.J. & Shams, L. (2015) Perception of body ownership is driven by Bayesian sensory inference. *PloS one*, **10**, e0117178.

Shams, L. & Beierholm, U.R. (2010) Causal inference in perception. *Trends Cogn Sci*, **14**, 425-432.

Shams, L., Kamitani, Y. & Shimojo, S. (2000) Illusions: What you see is what you hear. *Nature*, **408**, 788-788.

Shams, L., Kamitani, Y. & Shimojo, S. (2002) Visual illusion induced by sound. *Cognitive Brain Research*, **14**, 147-152.

Shams, L., Ma, W.J. & Beierholm, U. (2005) Sound-induced flash illusion as an optimal percept. *Neuroreport*, **16**, 1923-1927.

Author Manuscript

This article is protected by copyright. All rights reserved

Stecker, G.C. & Middlebrooks, J.C. (2003) Distributed coding of sound locations in the auditory cortex. *Biol Cybern*, **89**, 341-349.

Ursino, M., Cuppini, C. & Magosso, E. (2014) Neurocomputational approaches to modelling multisensory integration in the brain: a review. *Neural Netw*, **60**, 141-165.

Ursino, M., Cuppini, C. & Magosso, E. (2015) A neural network for learning the meaning of objects and words from a featural representation. *Neural Netw*, **63**, 234-253.

Ursino, M., Cuppini, C. & Magosso, E. (2017) Multisensory Bayesian Inference Depends on Synapse Maturation during Training: Theoretical Analysis and Neural Modeling Implementation. *Neural Comput*, **29**, 735-782.

Ursino, M., Magosso, E. & Cuppini, C. (2009) Recognition of Abstract Objects Via Neural Oscillators: Interaction Among Topological Organization, Associative Memory and Gamma Band Synchronization. *Neural Networks, IEEE Transactions on*, **20**, 316-335.

Wallace, M.T., Roberson, G.E., Hairston, W.D., Stein, B.E., Vaughan, J.W. & Schirillo, J.A. (2004) Unifying multisensory signals across time and space. *Exp Brain Res*, **158**, 252-258.

Wozny, D.R., Beierholm, U.R. & Shams, L. (2008) Human trimodal perception follows optimal statistical inference. *J Vis*, **8**, 24-24.

Wozny, D.R., Beierholm, U.R. & Shams, L. (2010) Probability matching as a computational strategy used in perception. *Plos Comput Biol*, **6**, e1000871.

Wozny, D.R. & Shams, L. (2011) Computational characterization of visually induced auditory spatial adaptation. *Front Integr Neurosci*, **5**, 75.

Yamashita, I., Katahira, K., Igarashi, Y., Okanoya, K. & Okada, M. (2013) Recurrent network for multisensory integration-identification of common sources of audiovisual stimuli. *Frontiers in computational neuroscience*, **7**.

Author Manuscript

This article is protected by copyright. All rights reserved

Zhang, W.-h., Chen, A., Rasch, M.J. & Wu, S. (2016) Decentralized multisensory information integration in neural systems. *J Neurosci*, **36**, 532-547.

Author Manuscript

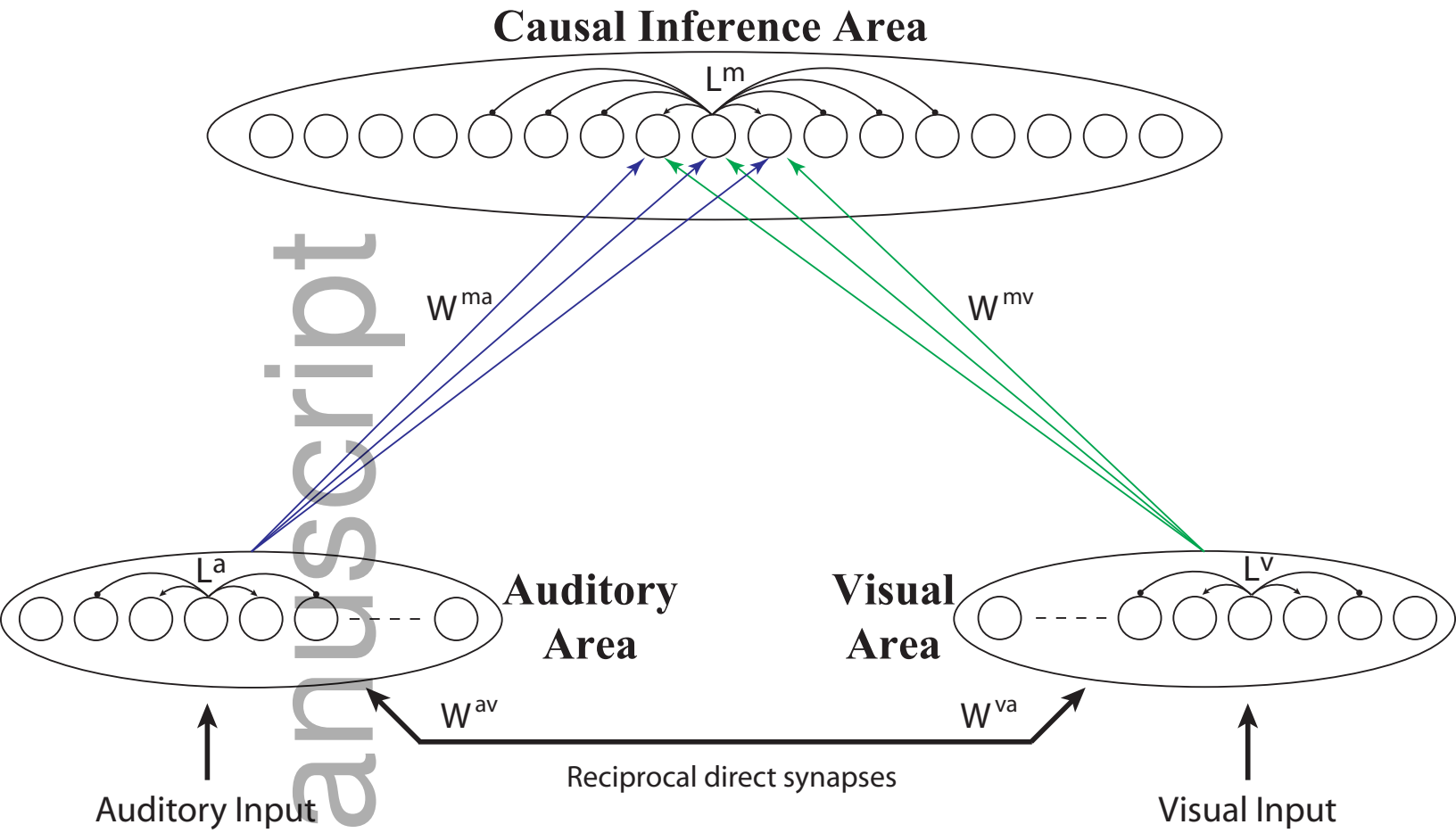
This article is protected by copyright. All rights reserved

Table I – Variance of Sensory Perception

	Auditory Area	Visual Area	Multisensory Area
Auditory Input	6.3768	/	6.6330
Visual Input	/	1.0347	1.2506
Visual-Auditory Input	2.5251	0.9416	1.3500

Author Manuscript

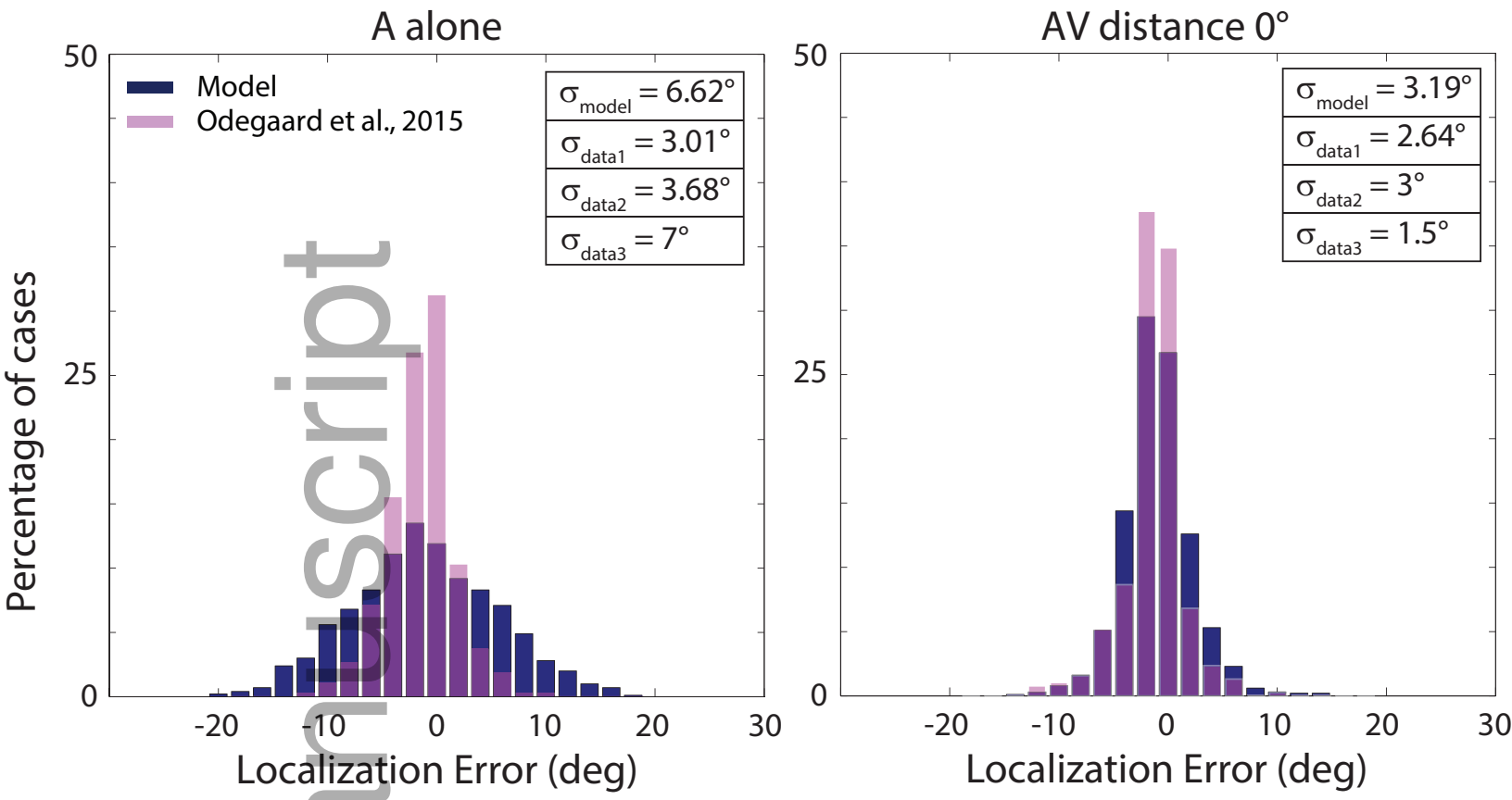
This article is protected by copyright. All rights reserved



ejn_13725_f1.eps

Author Manuscript

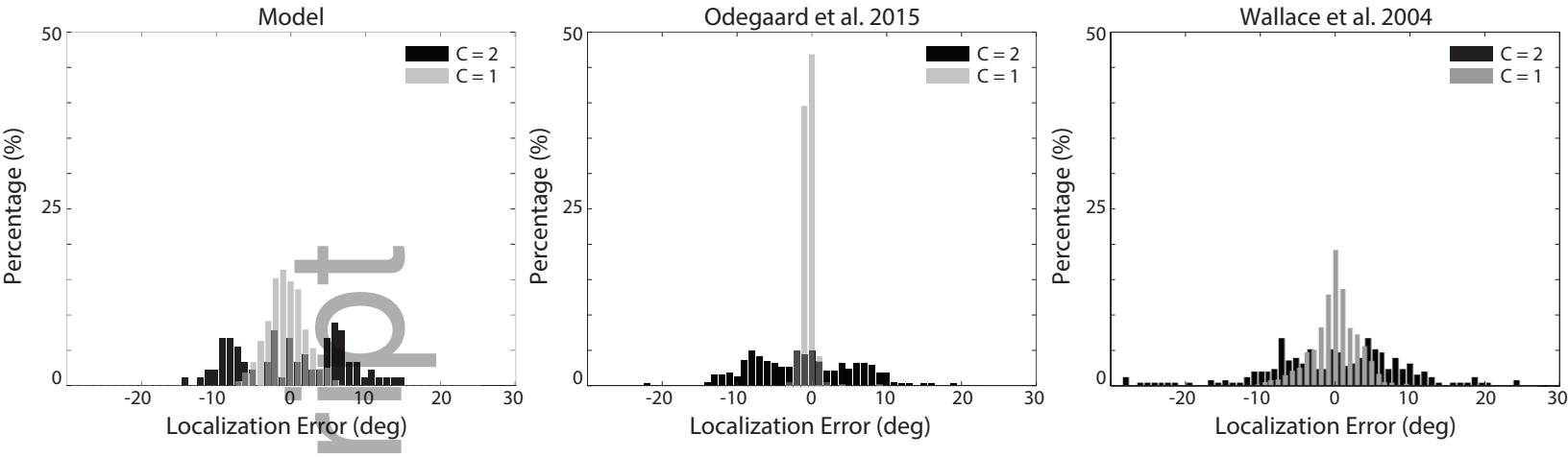
This article is protected by copyright. All rights reserved



ejn_13725_f2.eps

Author Manuscript

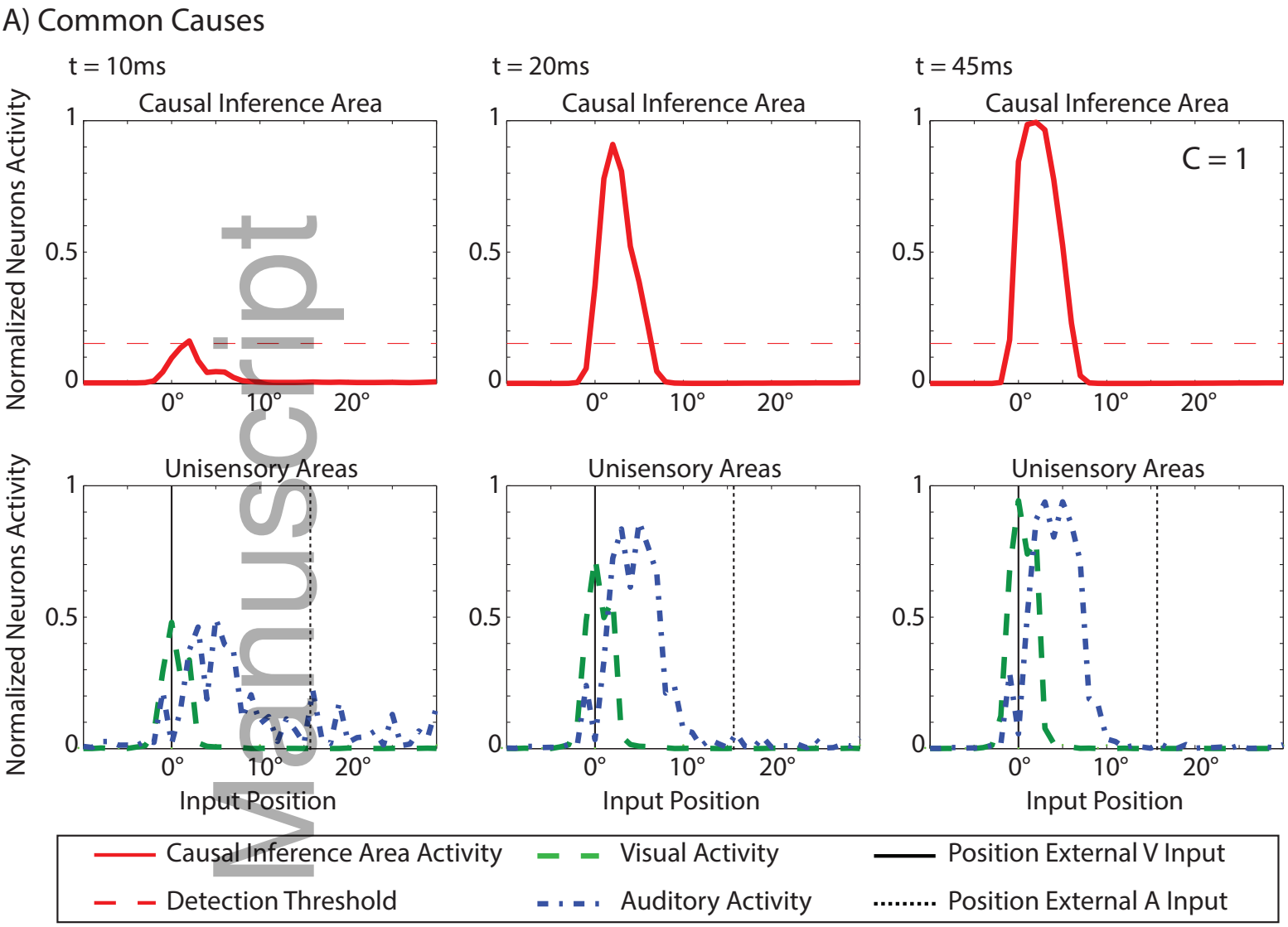
This article is protected by copyright. All rights reserved

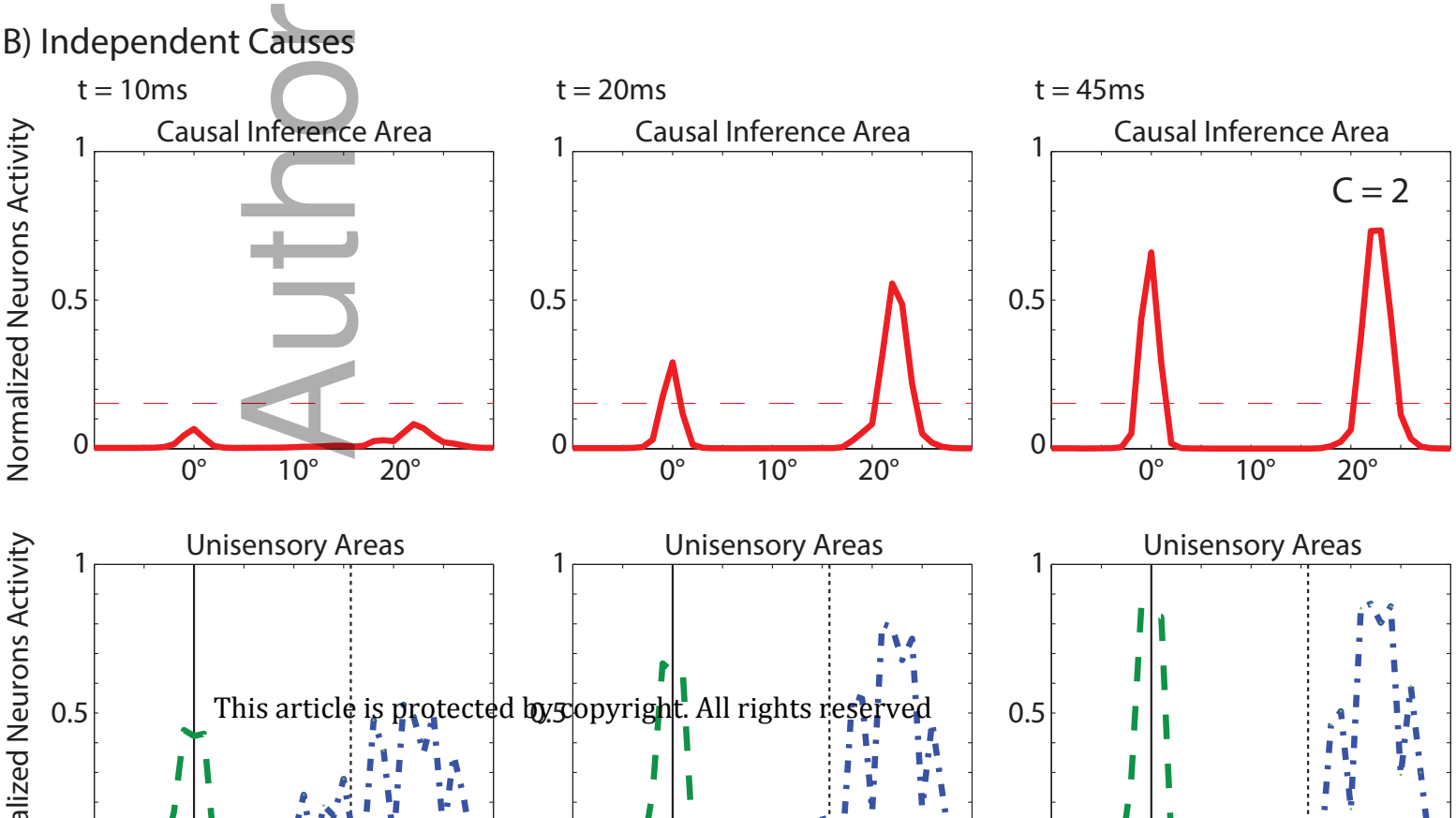


ejn_13725_f3.eps

Author Manuscript

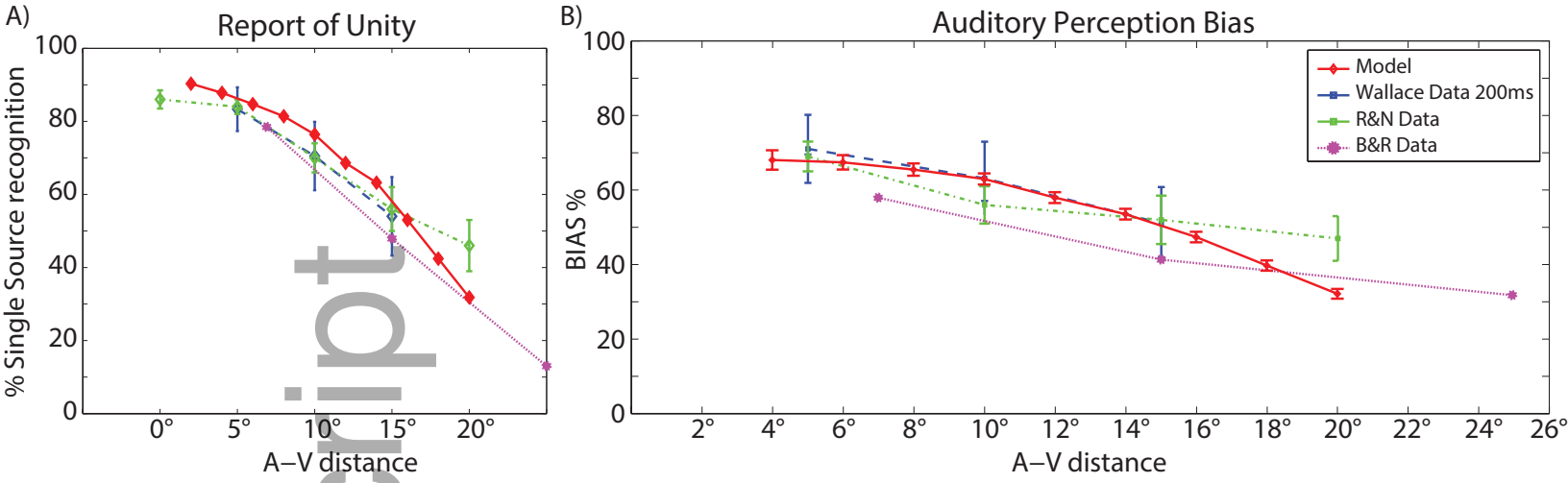
This article is protected by copyright. All rights reserved





Author Manuscript

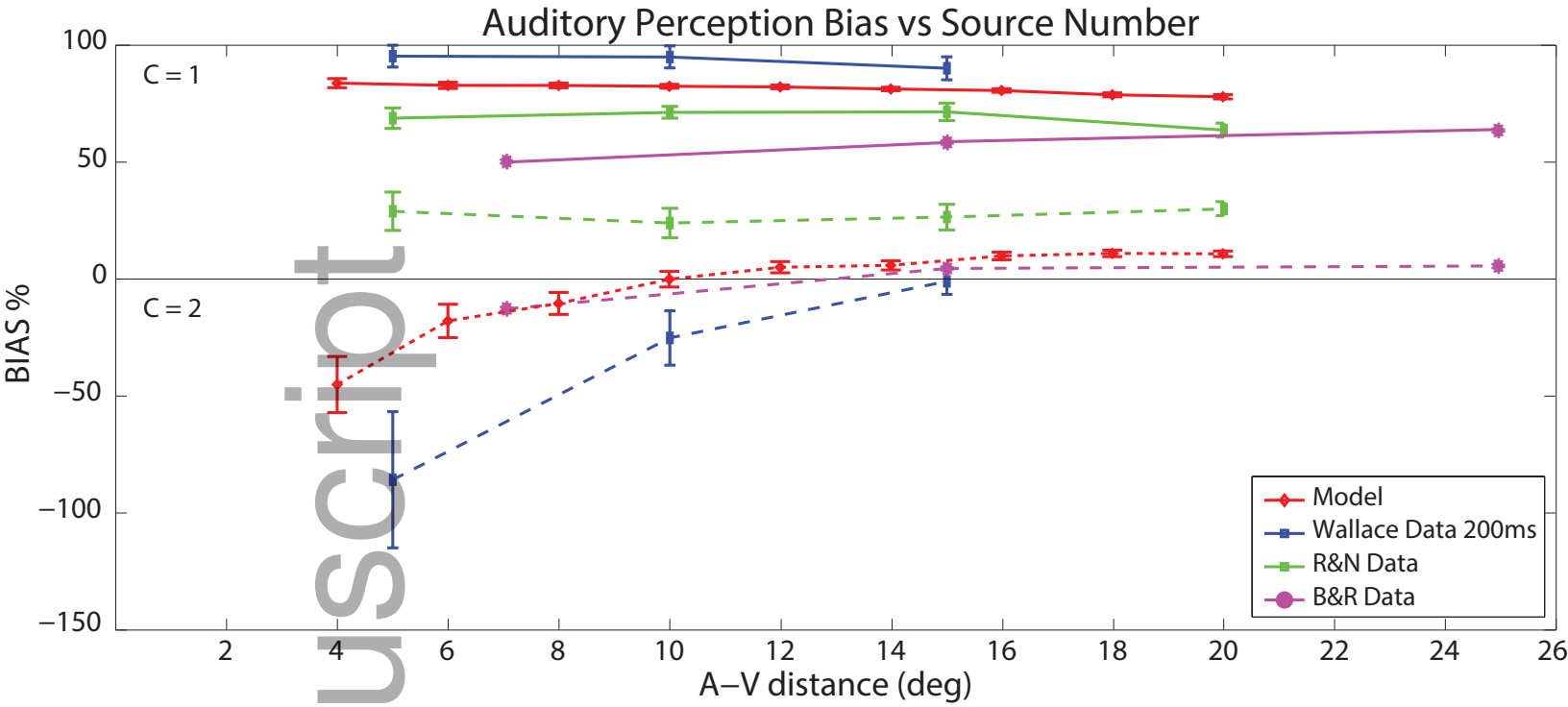
This article is protected by copyright. All rights reserved



ejn_13725_f5.eps

Author Manuscript

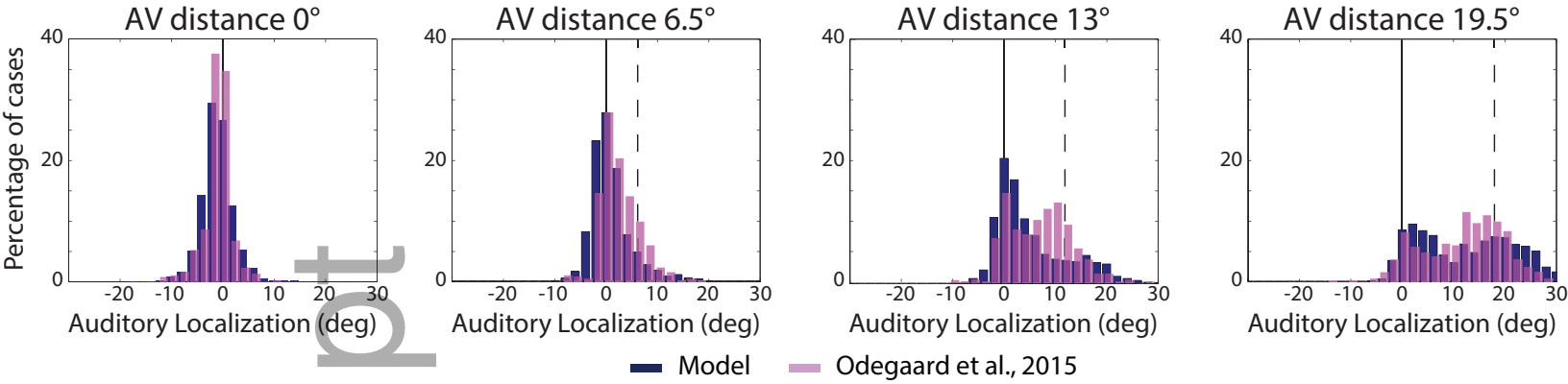
This article is protected by copyright. All rights reserved



ejn_13725_f6.eps

Author Manuscript

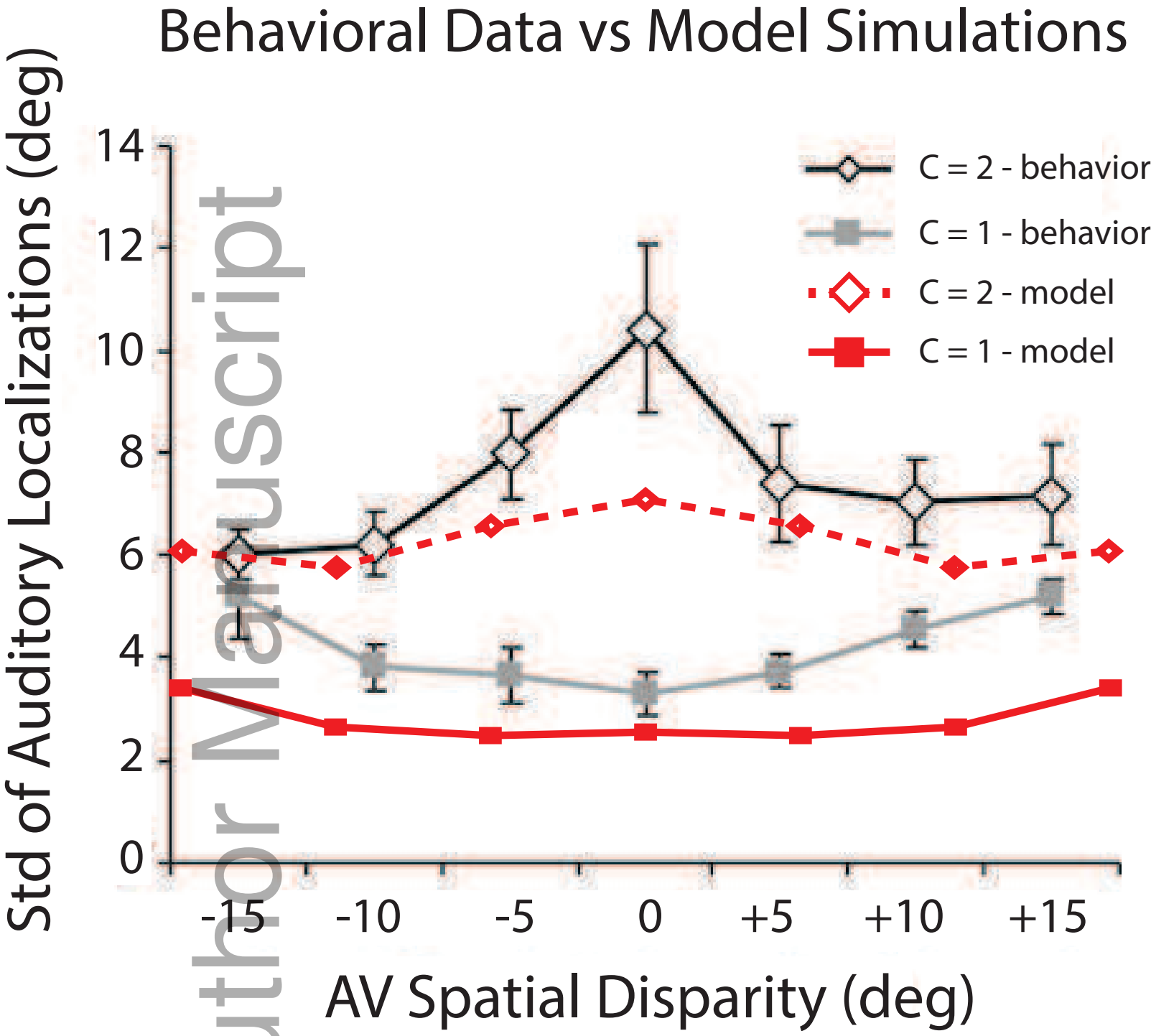
This article is protected by copyright. All rights reserved



ejn_13725_f7.eps

Author Manuscript

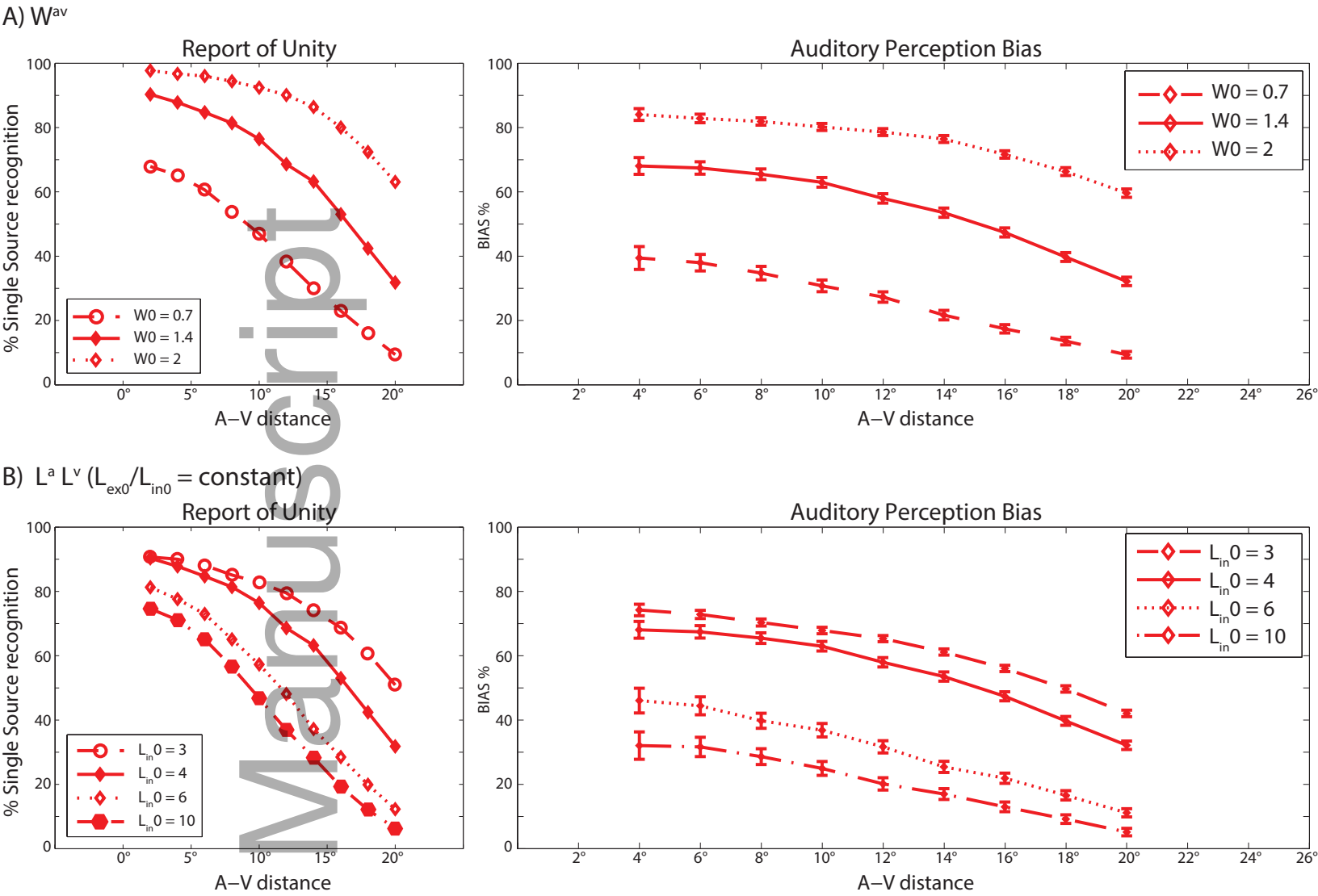
This article is protected by copyright. All rights reserved



ejn_13725_f8.eps

Author Manuscript

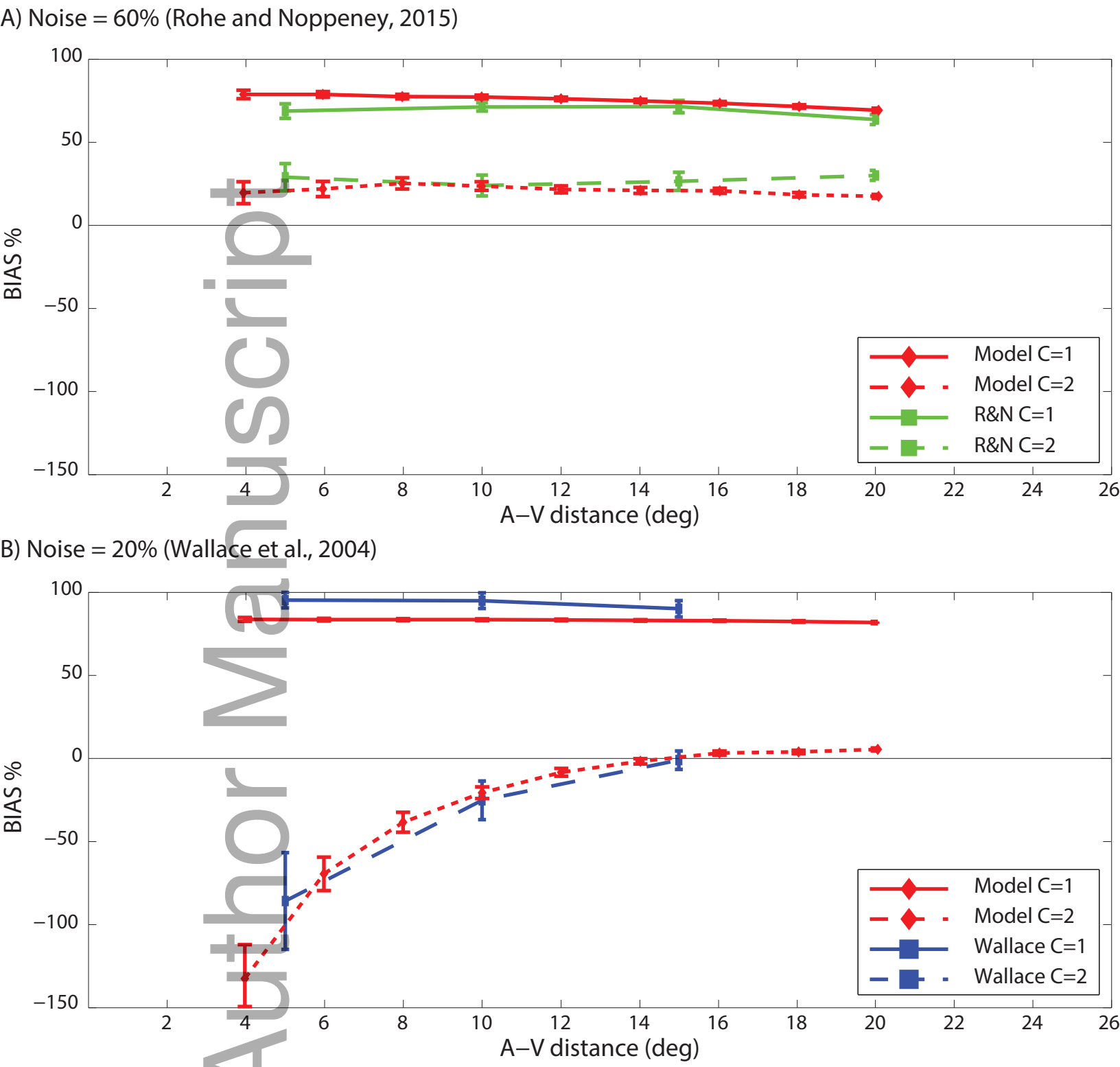
This article is protected by copyright. All rights reserved



ejn_13725_f9.eps

Author Manuscript

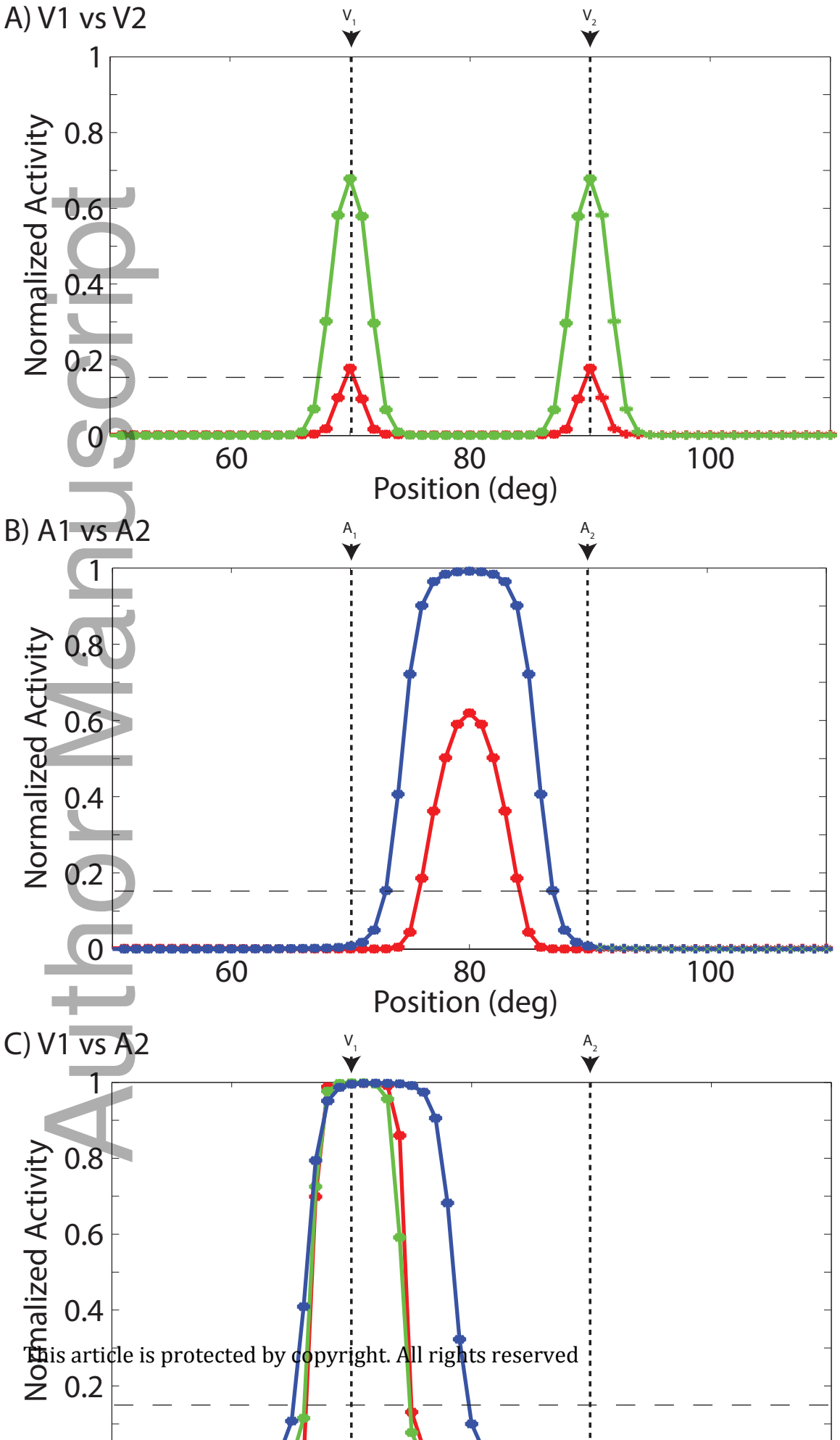
This article is protected by copyright. All rights reserved



ejn_13725_f10.eps

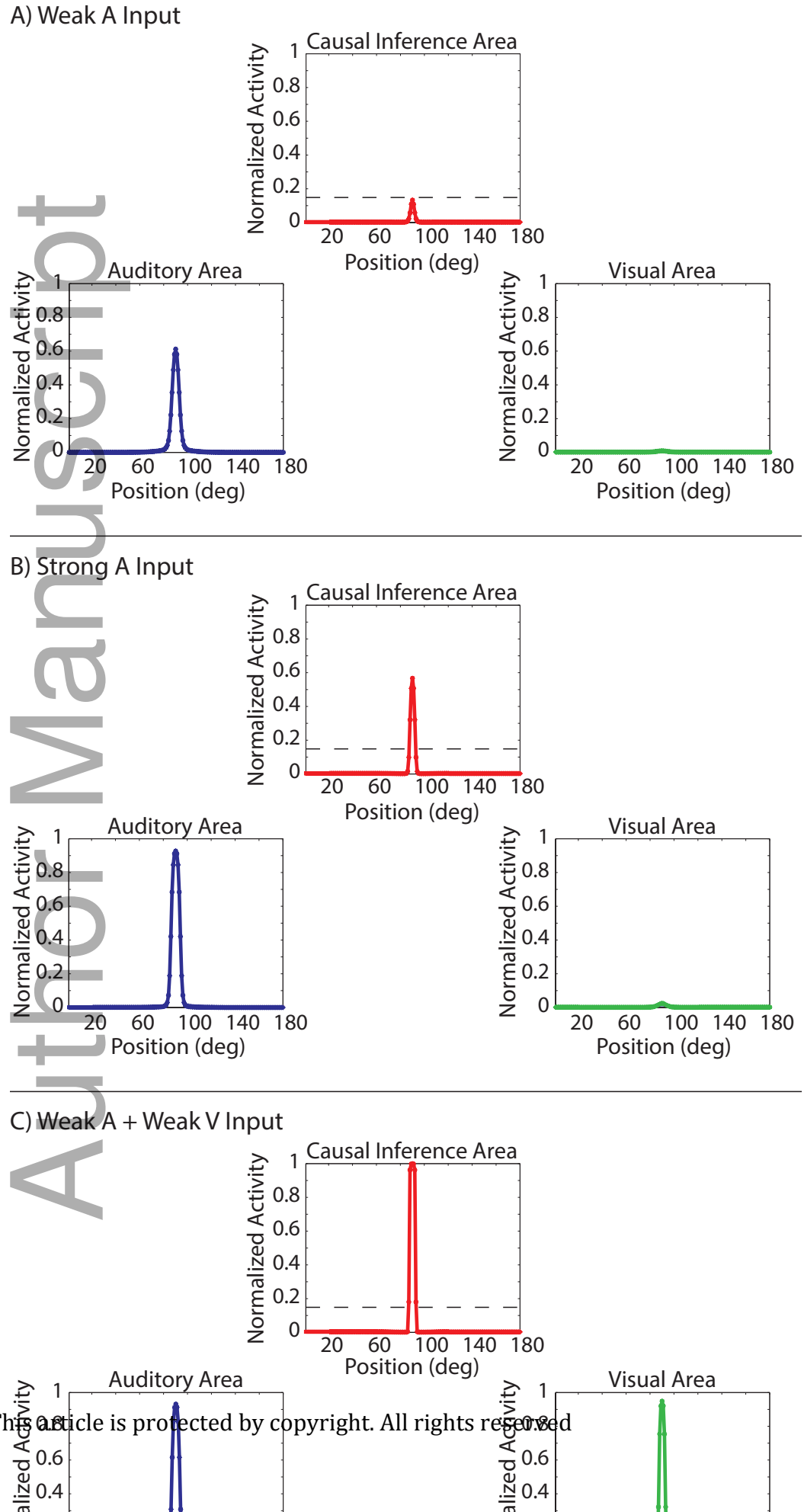
Author Manuscript

This article is protected by copyright. All rights reserved



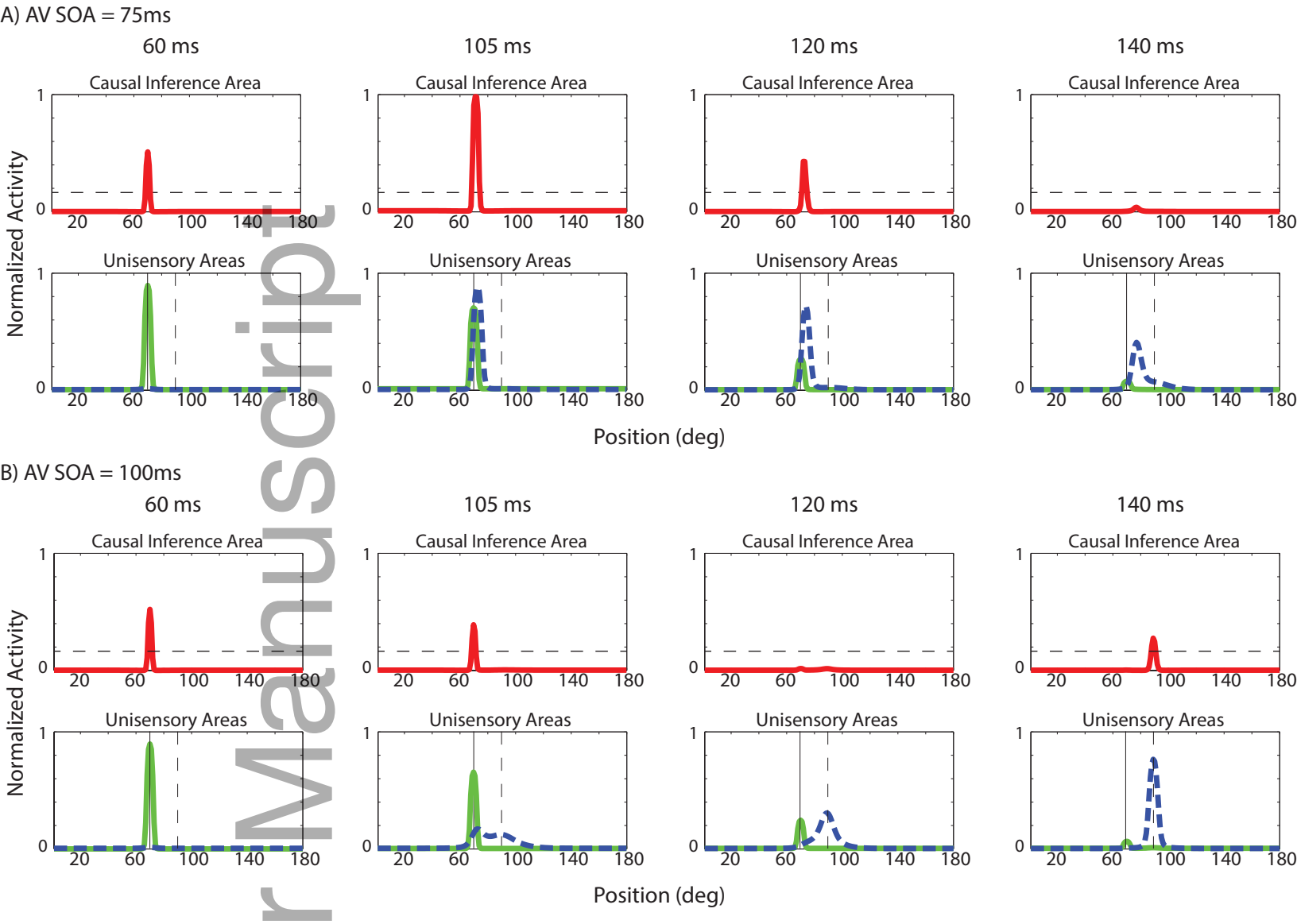
Author Manuscript

This article is protected by copyright. All rights reserved



Author Manuscript

This article is protected by copyright. All rights reserved



ejn_13725_f13.eps

Author Manuscript

This article is protected by copyright. All rights reserved