

# Hebbian Learning Mechanisms Help Explain the Maturation of Multisensory Speech Integration in Children with Autism Spectrum Disorder (ASD) and with Typical Development (TD): a Neurocomputational Analysis.

**Cristiano Cuppini (cristiano.cuppini@unibo.it)**

**Mauro Ursino (mauro.ursino@unibo.it)**

**Elisa Magosso (elisa.magosso@unibo.it)**

Department of Electric, Electronic and Information Engineering, University of Bologna, 2 Viale Risorgimento  
Bologna, 40136, Italy

**Lars A. Ross (lars.ross@einstein.yu.edu)**

**John J. Foxe (john.fox@einstein.yu.edu)**

**Sophie Molholm (sophie.molholm@einstein.yu.edu)**

Department of Pediatrics and Neuroscience, Albert Einstein College of Medicine, 1225 Morris Park Avenue  
Bronx, NY 10461, USA

## Abstract

Cognitive tasks such as communication and speech comprehension rely on the brain's ability to exploit and integrate sensory information of different modalities. Accordingly, the appropriate development of multisensory speech integration (MSI) greatly influences a child's ability to successfully relate with others. Several experimental findings have shown that speech intelligibility is affected by visualizing a speaker's articulations, and that MSI continues developing late into childhood. This work aims at developing a network to analyze the role of the sensory experience during the early stages of life, as a mechanism responsible for the maturation of these integrative abilities in teenagers. We extended a model realized to study multisensory integration in cortical regions (Magosso et al., 2012; Cuppini et al., 2014) by incorporating a multisensory area known to be involved in audiovisual speech processing, the superior temporal sulcus (STS). The model suggests that the maturation of MSI is primarily due to the maturation of direct connections among primary unisensory regions. This process was the results of a training phase during which the network was exposed to sensory-specific and cross-sensory stimuli, and excitatory projections among the unisensory regions of the model were subjected to Hebbian rules of potentiation and depression. With such a model, we also analyzed the acquisition of adult MSI abilities in ASD children, and we were able to explain the delayed maturation as result of a lower level of multisensory exposures during early phases of life.

**Keywords:** Autism Spectrum Disorder (ASD); Neural Networks; Hebbian Learning Rules; Multisensory Speech Integration; McGurk Effect

## Introduction

The brain's ability to exploit and integrate sensory information of different modalities is fundamental not just for simple detection tasks, but also for more demanding perceptual-cognitive functions, such as those involved in communication. For example, the intelligibility of speech is

significantly improved when one can see the speaker's articulations. Accordingly, the appropriate development of multisensory speech integration (MSI) greatly affects a child's ability to relate with others. Ample experimental evidence has shown that MSI appears to be highly immature at birth and that continues to develop late into childhood (Brandwein et al., 2010). Moreover, children with autism spectrum disorder (ASD) presenting impaired MSI early in life, show an amelioration in the adolescent years (de Boer-Schellekens et al., 2013; Foxe et al., 2015). These evidences suggest that there may be delays in the maturation of MSI for children with ASD that resolve at this point. Multiple studies have shown multisensory processing deficits in ASD in the absence of comparable unisensory deficits, suggesting that they represent impairment of neural processes that have direct and specific impact on MSI. However, the neural basis of the impairment remains unknown.

A region of particular interest for the maturation of MSI is the superior temporal sulcus (STS), an association cortex involved in speech perception (Molholm et al., 2013) that is also frequently implicated in audiovisual multisensory processing (Bolognini et al., 2009). This region must be considered in the context of its feedforward inputs from auditory and visual cortices. Converging evidence reveals that MSI occurs at very early stages of cortical processing and in sensory cortical regions, although the functional role of early MSI (at the onset of cortical sensory processing in some cases; Molholm et al., 2002) remains unknown.

Several experimental data pointed out that auditory speech recognition is relatively mature at 5 to 9 years of age, approaching adult-like performances (e.g., Fallon, Trehub & Schneider, 2000; Kraus, Koch, McGee, Nicol, & Cunningham, 1999), at ages where multisensory speech processing is not (Foxe et al., 2015).

Such observations suggest a neural model in which the maturation of MSI in speech perception follows from the reinforcement of direct “cross-modal” excitatory connections between auditory and visual speech representations in unisensory cortices. In this case, it can be assumed that connections among unisensory areas are initially relatively ineffective, but that they strengthen as a consequence of relevant multisensory experiences through a Hebbian learning mechanism. Thus, multisensory experiences would affect only the ability of STS elements to detect multisensory stimuli, via a reciprocal reinforcement of unisensory activities when both are active, but it would not produce any additional level of information to the STS in case of unisensory stimulation.

The aim of the present work is to test the feasibility of this model, and its consequences by using a computational model inspired by neurophysiology and based on a previous model implemented to study cortical multisensory interaction (Magosso et al., 2012; Cuppini et al., 2014). In particular, with the model we wish to i) analyze possible mechanisms underlying the maturation of MSI; ii) test the model's ability to reproduce different results concerning speech MSI in terms of accuracy as well whether it produces the well-known McGurk illusion; and iii) provide possible explanations of the neural processing differences that could lead to a slower maturation of MSI in participants with ASD, followed by a full recovery during adolescence.

In particular, we describe the training mechanisms implemented to simulate the maturation phase and we test a hypothesis to explain ASD deficits in speech MSI: a different multisensory experience during the maturation process, due to a lack of attention in young children (attentional bias) is responsible for the different maturation in ASD. All the simulated responses are compared with behavioral data present in the literature.

## Method

The model consists of a multisensory region (STS) of  $N$  multisensory units ( $N = 180$ ), receiving excitatory projections from two arrays of  $N$  auditory and  $N$  visual units (see Fig. 1). Unit response to any input is described with a first order differential equation, which simulates the integrative properties of the cellular membrane, and a steady-state sigmoidal relationship, that simulates the presence of a lower threshold and an upper saturation for neural activation. The saturation value is set at 1, i.e., all outputs are normalized to the maximum. In the following, the term “activity” is used to denote unit output.

**Auditory and visual units** are devoted to the processing of information regarding speech sounds and speech gestures (i.e. lip and face movements; see e.g., Bernstein & Liebenthal, 2014), and are topologically organized according to a similarity principle. This means that two similar sounds or lips movements activate proximal neural groups in these areas. The topological organization in these cortical regions is realized assuming that each unit is connected with other elements of the same area via lateral

excitatory and inhibitory connections (intra-area connections,  $L$  in Fig. 1), described by a Mexican hat disposition, i.e., proximal units excite reciprocally and inhibit more distal ones. This disposition produces an “activation bubble” in response to a specific auditory or visual input: not only the neural element representing that individual feature is activated, but also the proximal ones linked via sufficient lateral excitation. This arrangement can have important consequences for the correct perception of phonemes, for instance resulting in the illusory perceptual phenomena like the well-known McGurk effect (see section Results). In this work, lateral intra-area connections are not subject to training, since we assumed that this process took place earlier in life than the acquisition of MSI.

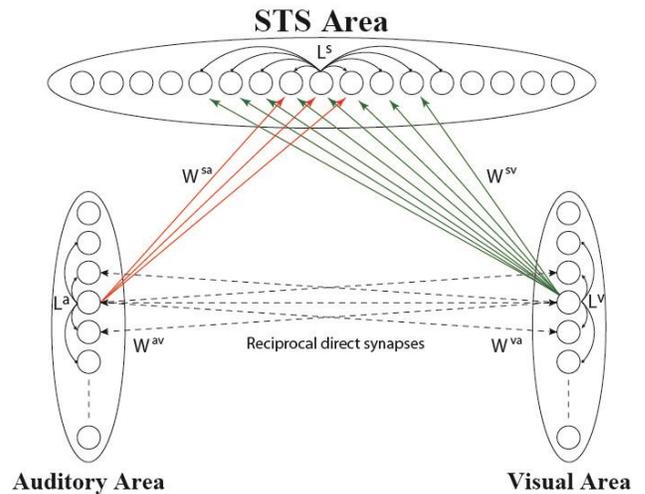


Figure 1: Architecture of the network. Each circle represents a unit. Each region is made of 180 elements. Dashed lines represent weights ( $W^{av}$ ,  $W^{va}$ ) acquired during a crossmodal training, which simulates associative learning between speech sound and gestures. Units in the same region are reciprocally connected through lateral synapses ( $L^a$ ,  $L^v$  and  $L^s$ ), described by a Mexican Hat function. Units in the unisensory regions send excitatory connections ( $W^{sv}$ ,  $W^{sa}$ ) to the corresponding elements in the multisensory area.

Furthermore, units in the auditory and visual regions also receive an external input (corresponding to a speech sound and/or a gesture representation of the presented phoneme). These visual and auditory inputs are described with a gaussian function. The central point of the Gaussian function corresponds to a specific speech sound/gesture, and its amplitude with the stimulus intensity; the standard deviation accounts for the uncertainty of the stimulus representation. In this model, for simplicity the two inputs are described with the same function. To reproduce experimental variability, the external input had been added with a noisy component, taken from a uniform distribution. Moreover, since the outside inputs are mediated by long-range excitatory connections, their temporal aspects are described by using a second order kinetics, similar to that commonly adopted to mimic the glutamatergic synaptic

response (i.e., an impulse produces a response similar to an alpha function, see also Jansen and Rit, 1995). These kinetics are characterized by different time constants ( $\tau^a$  and  $\tau^v$  for the two modalities) simulating the auditory and visual processing in the cortex.

Finally, we consider a cross-modal input, computed assuming that units of the two areas could be reciprocally linked via long-range excitatory connections ( $\mathbf{W}^{av}$ ,  $\mathbf{W}^{va}$ , in Fig 1), described by a pure latency, and the same second-order kinetics employed to mimic the temporal aspects of the external inputs. We assume that, in the network's initial configuration, corresponding to an early period of life, these connections have negligible strength, but are subject to a training phase (see below) during which the network learns to associate the auditory (speech sounds) and visual (speech gestures) representations of the same phonemes.

The third area simulates **multisensory units** in a cortical region (STS) known to be involved in the phoneme comprehension tasks, and MSI. These units are linked via lateral connections with a Mexican-hat arrangement, implementing a similarity principle ( $L^s$ , in Fig. 1).

Inputs to the multisensory area were generated by long-range excitatory connections from unisensory regions ( $\mathbf{W}^{sv}$ ,  $\mathbf{W}^{sa}$ ): we used a delayed onset (pure latency) and a second-order kinetics to mimic the temporal aspects of these inputs. The connections between unisensory and multisensory regions were realized with a Gaussian function, assuming stronger and more focused connections coming from the auditory region ( $\mathbf{W}^{sa}$ ), and more diffuse but weaker connections coming from the visual area ( $\mathbf{W}^{sv}$ ). This asymmetric connectivity helps explain the experimental results present in the literature about the better abilities in speech identification in case of auditory stimulation, compared with the poor performance in the case of visual inputs. This different representation is assumed being the final state of a process of unisensory maps refinement in STS, which takes place in early stages of life. This development could be included in future implementations of the model, as an earlier training phase based on the evidence that auditory stimuli are more informative than the visual representations of words. The feedforward connectivity is also responsible for the presence of early weak integrative phenomena in the younger ASD group (Fig. 1A, Foxe et al.). In the present model, neither of these connections are modified during the learning period due to the relative stability of representations of unisensory speech features at the ages considered (~7 years of age upward). Finally, the output of the STS units is compared with a fixed threshold to mimic the perceptual ability to correctly identify speech (detection threshold).

### Training the Network

We simulated a normal training period by presenting thousands (up to 25.000 inputs) of unisensory and multisensory speech representations to the network, to mimic a normal experience with speech stimuli: specifically we trained the network with 80% of congruent auditory and

visual stimuli and 20% of auditory stimuli alone. During the training phase we used suprathreshold stimuli at their highest level of efficacy, i.e. stimuli able to excite unisensory units close to the upper saturation, in order to speed up the modeling process.

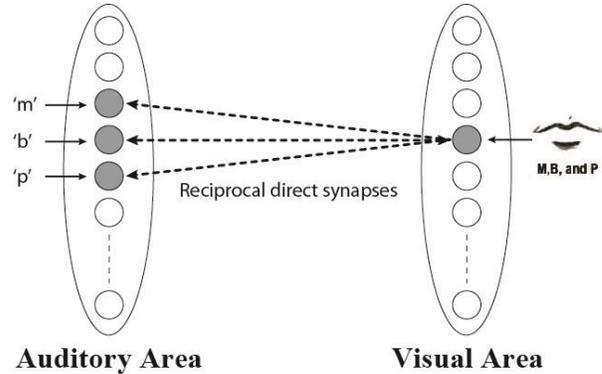


Figure 2: Representation of training mechanisms simulating the associative learning between sounds (Auditory area) and gestures (Visual area) of units of speech. In case of multisensory stimuli, speech sounds are presented along with corresponding lip movements. Thanks to the Hebbian learning rules, connections among contemporarily active units are reinforced. Hence, the network learns how to associate the auditory and visual representations of the same speech events, and this knowledge is implemented in the synaptic architecture between the unisensory regions.

These stimuli were generated through a uniform distribution of probability. Each stimulus lasted 130 ms, during which, after an initial transient period, the connections among visual and auditory representations of the same phonemes were crafted by using Hebbian algorithms of long-term potentiation (LTP) and long-term depression (LTD). In particular, we chose a presynaptic gating rule, which means that the training algorithm only modifies the connections coming from an active unit, and their strength is modified based on the activity of the postsynaptic units. As an example, if a presynaptic auditory element is active, it reinforces connections targeting a simultaneously active visual unit (likely representing the same speech unit), and weakens connections with silent visual elements (likely those coding for different speech inputs, see Fig. 2) (see Gerstner, W., & Kistler, 2002). In order to establish this correlation, the activity of the individual units (both presynaptic and postsynaptic) is compared with a given threshold, to determine whether the unit can be considered active or silent. The strengthening and depression processes are also subject to a saturation rule: which means that each single connection cannot overcome a maximum value, nor decrease below zero.

Finally, to simulate the delayed developmental processes taking place in ASD children, we trained and tested the network by using lower multisensory experiences, precisely 20% multisensory stimuli plus 80% auditory stimuli.

## Results

A first set of simulations was performed to evaluate the network’s ability to correctly identify speech before the model had been exposed to training (Fig. 3). Already mature unisensory maps in auditory and visual regions were supposed in this model, as described in previous section.

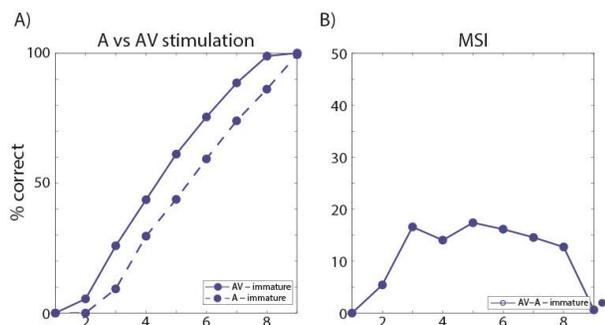


Figure 3: Average word recognition performance (% correct) before training. Panel A reports the percentage correct speech recognition (y-axes) in case of auditory stimulation (dashed line), or multisensory stimulation (solid line). These data represent the mean of correct recognitions over 3600 different presentations for each level of stimulus efficacy (reported in the x-axes). A correct recognition has been computed comparing the activity elicited in a unit in the STS region, coding for a specific phoneme, with a threshold (fixed at the 30% of its maximum value). Panel B reports the Multisensory Speech Integration (MSI) abilities of the network, computed as the difference between the percentage of correct detections in case of crossmodal stimulations and its counterpart in case of auditory stimuli.

In this phase, representations of speech in the two unisensory regions are independently activated by the two modality-specific external stimuli, and do not interact through direct long-range excitatory projections between the unisensory cortical regions, which are still ineffective. Hence, they independently stimulate the corresponding units in STS region. As shown in Fig. 3, in this initial condition, an effective auditory stimulus alone is sufficient to produce a high percentage of correct speech sound identifications, as in mature adult-like behaviour. If the auditory stimulus is coupled with a simultaneous visual representation of the same phoneme, the network shows some benefit, although this is relatively low and no greater than 20% MSI gain over all stimuli and levels of efficacy.

So, the network in its initial stage is characterized by: i) mature abilities in speech-recognition tasks in case of auditory-alone stimulation, but ii) poor multisensory integration (see Fig. 3). These results are in agreement with what one would expect prior to significant training, and indeed are well aligned with what we see in our data in which younger children show relatively immature ability to benefit from MSI, whereas auditory speech recognition is significantly closer to mature performance levels.

## Developmental process and audio-visual speech recognition

The model in its initial state was repeatedly stimulated with modality-specific and cross-modal inputs (see section Training) in order to simulate the experience of a child with different sensory representations of phonemes. The weights of the inter-area projections among unisensory elements in the visual and auditory regions adjusted according to Hebbian dynamics. We tested MSI in the final “adult-like” configuration and throughout the developmental process, using the same testing paradigm used to evaluate the MSI behavior in the immature phase.

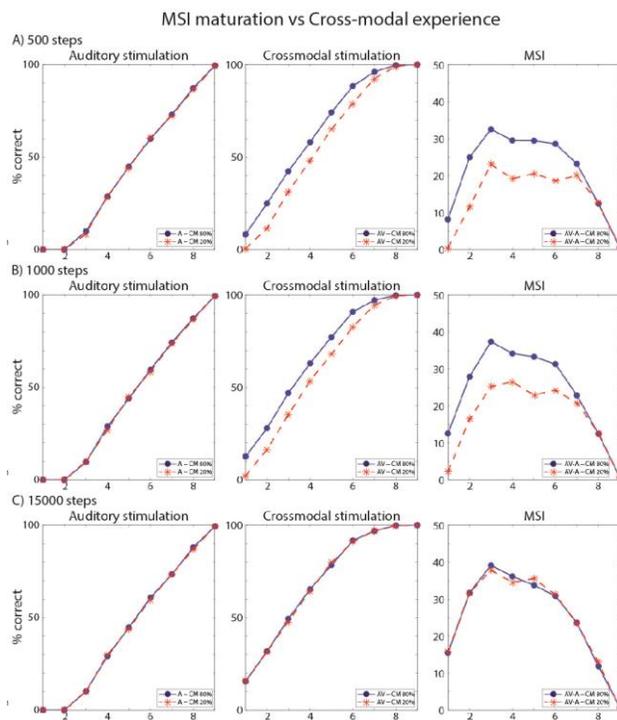


Figure 4: MSI acquisition with a multisensory experience of 20% during the training (red lines) compared with the normal multisensory experience (blue lines).

One possible explanation for reduced MSI in ASD is that learning is less effective in this group. A possible explanation tested here is that these individuals experience fewer multisensory exposures, possibly due to how attention is allocated (e.g., suppression of unattended signals; selectively focusing on one sensory modality at a time; not looking at faces consistently). We therefore tested the impact of percentage of multisensory versus unisensory exposures on model performance on the maturation of MSI.

Fig. 4 reports the weight maturation (left panel) and MSI abilities (right panel) at different epochs, for a training phase in which the network was exposed to a sensory training with just 20% of multisensory stimuli.

Even with such a poor multisensory experience, the network can reach “TD-like” behaviour in terms of MSI as shown in Fig. 4, although this maturation requires 15,000

training epochs. This result suggests that multisensory integration in the model strongly depends on connections from the visual to the auditory region.

### **Simulation of the McGurk effect**

An important consequence of training in our model is that the audio-visual inference becomes stronger after training, because of connection-weight reinforcement among unisensory areas. This change has important consequences in the development of audio-visual illusions. Since unisensory areas in our model code for speech, a typical illusion consists in the well-known McGurk effect (McGurk & McDonald, 1976). In this illusion, incongruent auditory speech is dubbed onto visual speech and the resulting auditory speech percept corresponds to a fusion of the auditory and visual speech stimuli, or to the visual speech stimulus, but not to the veridical auditory speech stimulus.

We performed an additional set of simulations with the model (both in the mature and immature configurations) to reproduce a McGurk-type situation. Specifically, we presented mismatched (at four-position distance) auditory-visual speech to the network and analyzed the activities elicited in all areas. We say that the McGurk effect is evident when the detected phoneme (computed as the barycenter of activity in the multisensory region) is different from that used in the auditory input. The network in the immature configuration is characterized by limited visual influence on the speech percept. Therefore, the activity in the auditory region is almost unaffected by the visual stimulus. In this case, the auditory modality plays the dominant role in guiding speech perception. In the 42.5% of presentations, the model identifies the auditory input correctly, while the McGurk effect is present less than 30% of the time. In the remaining 27.2% of cases, no phoneme reaches the detection threshold.

After training, the model is much more susceptible to the AV illusion, with responses affected by the visual information on almost 72% of the simulation trials.

### **Discussion**

Different computational models have been developed in recent years to investigate the general problem of multisensory integration in the brain (see Cuppini et al., 2011 and Ursino et al., 2014 as a review). Some of them, in agreement with several psychophysical and behavioral data, are based on a Bayesian approach (Anastasio et al., 2000; Knill and Pouget, 2004; Körding et al., 2007). Others assume that integration is an emergent property based on network dynamics (Patton and Anastasio, 2003; Ursino et al., 2009). Finally some models have been realized to deal with the problem of multisensory integration in semantic memory and lexical aspects (Rogers et al., 2004; Ursino et al., 2010, 2015).

Concerning the specific problem of speech recognition, Ma et al (Ma et al., 2009) implemented a Bayesian model of optimal cue integration able to explain visual influence on auditory perception in a noisy environment. They explained

different perceptual behaviors based on words representation as a collection of phonetic features in a topographically organized feature space.

Although the previous computational efforts simulated experimental data quite well, none of them was able to explain the maturation of MSI in speech perception or the different developmental trajectory for ASD, or how these capabilities are instantiated in the circuit.

The present model, in its mature architecture, simulates many experimental findings present in literature regarding speech MSI. From this point of view, the fundamental assumption is that the adult configuration implements a two-step cross-modal integration: the first at the level of unisensory areas, mediated by the cross-modal connections between visual and auditory regions; the second at the level of the multimodal area, due to the presence of convergent feedforward connections. With this model, we reproduced the improvement in correct phoneme recognition in audiovisual vs auditory conditions at different signal-to-noise levels (Fuxe et al., 2015); and, we simulated the main aspects of the McGurk effect.

A second important aspect of our study is the capacity to mimic and to understand the developmental differences between TD subjects and ASD children regard the cross-modal abilities observed with age. In particular, the model explains results of a recent study by Fuxe et al. (2015), using two main assumptions. First, the feedforward connections from unisensory areas to the multisensory area are already mature in the early age (here, this corresponds to the condition of the untrained network) and the auditory feedforward connections are stronger than the visual ones. Second, the cross-modal connections between unimodal areas are created during the development, under the pressure of a multimodal environment (i.e., auditory + visual stimuli) and this process is faster in TD subjects than in ASDs. This assumption agrees with the diffuse idea that ASD subjects have a decreased long-range connectivity, and that autism is a functional disconnection syndrome, in which the core of deficit derives from the poor capacity to functionally connect remote regions of the brain (Melillo & Leisman, 2009). Since the reason for this decreased connectivity is still unclear, the model tested a possible scenario where a reduced number of cross-modal stimuli (reflecting a reduced attention of the subject to the external world), is a likely mechanism responsible for the differences in TD and ASD.

These differences may lead to some testable predictions: from the results about the training phase, when can expect that ASD children trained with a high percentage of cross-modal stimuli, could exhibit a normal or at least a quicker MSI maturation. A second prediction is that as a consequence of poor cross-modal connections among unisensory areas, young individual with ASDs have a less evident McGurk effect, but at the end of the developmental phase, this illusion becomes comparable in the two classes. The first prediction is still to be tested; the second is supported by some experimental results in the literature, in particular by comparing data across different studies.

However, it deserves a deeper investigation through a single, ideally longitudinal study.

Future developments of this model may include a more detailed and biologically realistic description of the unisensory areas, and the inclusion of further regions to simulate the role in MSI and speech perception played by subcortical structures, like the thalamus and basal ganglia.

## References

- Anastasio, T.J., Patton, P.E., & Belkacem-Boussaid, K. (2000). Using Bayes rule to model multisensory enhancement in the superior colliculus. *Neural Computation*, 12, 1165–1187
- Bernstein, L.E., & Liebhenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in neuroscience*, 8.
- de Boer-Schellekens, L., Keetels, M., Eussen, M., & Vroomen, J. (2013). No evidence for impaired multisensory integration of low-level audiovisual stimuli in adolescents and young adults with autism spectrum disorders. *Neuropsychologia*, 51(14), 3004-3013.
- Bolognini, N., Miniussi, C., Savazzi, S., Bricolo, E., & Maravita, A. (2009). TMS modulation of visual and auditory processing in the posterior parietal cortex. *Experimental brain research*, 195(4), 509-517.
- Brandwein, A.B., Foxe, J.J., Russo, N.N., Altschuler, T.S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex*, 21(5), 1042-1055.
- 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. (2011). Organization, maturation, and plasticity of multisensory integration: insights from computational modeling studies. *Frontiers in psychology*, 2.
- Fallon, M., Trehub, S.E., & Schneider, B.A. (2000). Children's perception of speech in multitalker babble. *The Journal of the Acoustical Society of America*, 108(6), 3023-3029.
- Foxe, J.J., Molholm, S., Del Bene, V.A., Frey, H. P., Russo, N.N., Blanco, D., Saint-Amour, D., & Ross, L.A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, 25(2), 298-312.
- Gerstner, W., & Kistler, W.M. (2002). Mathematical formulations of Hebbian learning. *Biological cybernetics*, 87(5-6), 404-415.
- Jansen, B.H., & Rit, V.G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4), 357-366.
- Knill, D.C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27 (12), 712–719.
- Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE* 2 (9), e943.
- Kraus, N., Koch, D.B., McGee, T.J., Nicol, T.G., & Cunningham, J. (1999). Speech-Sound Discrimination in School-Age Children. Psychophysical and Neurophysiologic Measures. *Journal of Speech, Language, and Hearing Research*, 42(5), 1042-1060.
- Ma, W.J., Zhou, X., Ross, L.A., Foxe, J.J., & Parra, L.C. (2009). Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One*, 4(3), e4638.
- Magosso, E., Cuppini, C., & Ursino, M. (2012). A neural network model of ventriloquism effect and aftereffect. *PLoS one*, 7(8), e42503.
- McGurk H, MacDonald J. (1976). Hearing lips and seeing voices. *Science*, 264, 746-8.
- Melillo, R., & Leisman, G. (2009). Autistic spectrum disorders as functional disconnection syndrome. *Reviews in the Neurosciences*, 20(2), 111-132.
- Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive brain research*, 14(1), 115-128.
- Molholm, S., Mercier, M.R., Liebhenthal, E., Schwartz, T.H., Ritter, W., Foxe, J.J., & De Sanctis, P. (2014). Mapping phonemic processing zones along human perisylvian cortex: an electro-corticographic investigation. *Brain Structure and Function*, 219, 1369-1383.
- Nath, A.R., & Beauchamp, M.S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage*, 59, 781-787.
- Patton, P.E., & Anastasio, T.J. (2003). Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons. *Neural computation*, 15, 783-810.
- Rogers, T.T., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R., and Patterson, K. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111, 205-235.
- Ursino, M., Cuppini, C., & Magosso, E. (2010). A computational model of the lexical semantic system based on a grounded cognition approach. *Frontiers in Psychology*, 1, 221.
- Ursino, M., Cuppini, C., & Magosso, E. (2014). Neurocomputational approaches to modelling multisensory integration in the brain: A review. *Neural Networks*, 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 Networks*, 63, 234-253.
- Ursino, M., Cuppini, C., Magosso, E., Serino, A., & Di Pellegrino, G. (2009). Multisensory integration in the superior colliculus: a neural network model. *Journal of computational neuroscience*, 26(1), 55-73.