Gain-Field Modulation Mechanism in Multimodal Networks for Spatial Perception

Alexandre Pitti, Arnaud Blanchard, Matthieu Cardinaux, Philippe Gaussier

2012 12th IEEE-RAS International Conference on Humanoid Robots
Nov.29-Dec.1, 2012. Business Innovation Center Osaka, Japan

Gain-Field Modulation Mechanism in Multimodal Networks for Spatial Perception

Alexandre Pitti, Arnaud Blanchard, Matthieu Cardinaux, Philippe Gaussier
ETIS Laboratory, UMR CNRS 8051, the University of Cergy-Pontoise, ENSEA, France. Email: alexandre.pitti@ensea.fr

***Abstract*—Seeing is not just done through the eyes, it involves the integration of other modalities such as auditory, proprioceptive and tactile information, to locate objets, persons and also the limbs. We hypothesize that the neural mechanism of gain-field modulation, which is found to process coordinate transform between modalities in the superior colliculus and in the parietal area, plays a key role to build such unified perceptual world. In experiments with a head-neck-eye's robot with a camera and microphones, we study how gain-field modulation in neural networks can serve for transcribing one modality's reference frame into another one (e.g., audio signals into eyes' coordinate). It follows that each modality influences the estimations of the position of a stimulus (multimodal enhancement). This can be used in example for mapping sound signals into retina coordinates for audio-visual speech perception.**

I. Introduction

Perceiving objects in space is one of first tasks babies have to deal with during infancy. It is a rather difficult problem since infants have to represent one object with multiple sensory modalities (vision, sound, tactile) encoded in different reference frames (e.g., eye-centered, head-centered or hand-centered). This curse of dimensionality corresponds to the so-called binding problem across the modalities for which there is still debates on its underlying neural mechanisms and its associated computational models. In this paper, we propose to inspire ourself from current computational models of spatial cognition and to investigate a neural system that potentially satisfies developmental evidences and biological mechanisms.

A reason for the weakness of spatial cognition in infants is that motor development is not fully mature during the first year. Infants calibrate the basic sensory-motor relationships during motor babbling; e.g., he locates his hand with its eyes because he moves the hand only when it crosses the eye-field. By doing so, he estimates its position in space relative to his head for example, on a univocal reference frame [1], [2]. The superior colliculus plays an important role for building such spatial map [3], [4]. Its principal functions serve for simple orientation tasks toward a target, like guiding saccadic eye movements to the locations of both auditory and visual stimuli.

Similar properties are observed in the parietal cortex, which is also hypothesized to maintain a mapping of sensory coordinates of objects into motor coordinates [5], [6].

From a biological viewpoint, it is interesting to note that both structures, the superior colliculus and the parietal cortex, which are found important for spatial cognition and multimodal integration, rely on the two same neural mechanisms, namely *(1)* timing integration and *(2)* gain-field modulation [7]. The first mechanism, the timing integration in sensory-motor circuits, means the detection of temporal events like synchrony or rhythmicity, it leads the perceptual enhancement or discrepancy in attentional tasks by reinforcing the links of contingent neurons, as emphasized in Hebb's law [8]. The second mechanism, the gain-field modulation, describes the phenomenon where the motor and the sensor signals mutually influence the amplitude activity of their afferent parietal neurons [9]. Differently said, these neurons encode stimulus location simultaneously in more than one reference frame using "gain fields" [5], [7]. For instance, there is a non-linear dependency on eye position for certain visual neurons in posterior parietal neurons (PNNs) whose reference frame is centered on the head, whilst others are found to be influenced more by the coding of somatic information into hand/arm-centered reference frame.

Gain modulation contributes therefore as a major computational mechanism for coordinates transformation and for the compensating of distortions caused by movements [10]. Its role is even broader as PNNs are found also important for reaching targets, goal-directed movements [11] and even for intentional acts [12]. Pouget and Deneve suggest that the parietal neurons behave as a population of basis functions that are continuously adapting their dynamics to the current coordinate frame relative to the task [5], [13]. All-in-all, these considerations suggest that exploiting the mechanisms of synchronization and of gain modulation may permit to understand how multimodal maps compute spatial processing.

Many computational frameworks have been proposed for multimodal integration, which are for some of them biologically-inspired. For instance, Fuke et al. [14] follow the models exploiting hebbian learning, using self-organizing maps (SOMs) to model the ventral intra-parietal (VIP) neurons responsible for facial somatic-visual integration. In their simulation, the SOMs estimate the relative arm position with respect to the face for visuo-tactile face representation where the synaptic links of the most contiguous visual and tactile neurons are reinforced over time. Besides, Chinellato et al. [15] follow the model proposed by Pouget and Deneve [13], which exploits the gain-field mechanism for multimodal integration. In a computer simulation of an eye-hand system, they use radial basis function networks (RBFs) for visuomotor transformations, for gazing and for reaching

978-1-4673-1369-8/12/$31.00 ©2012 IEEE

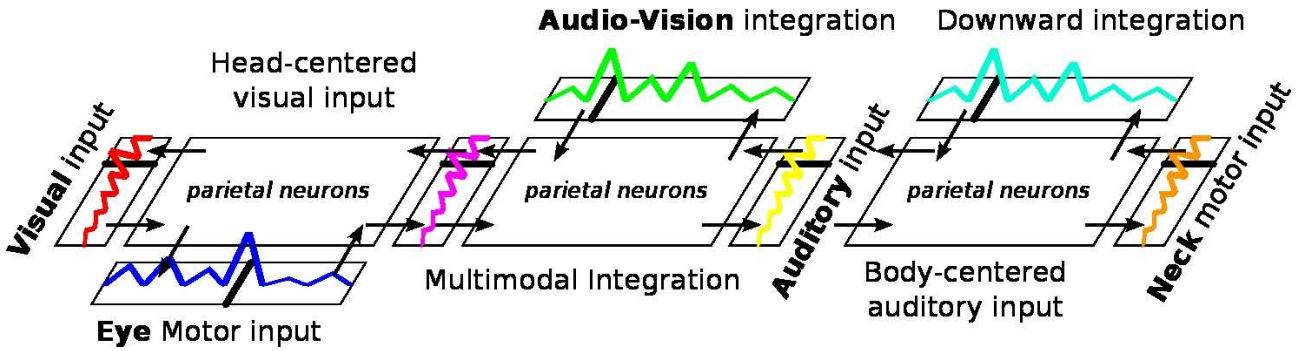


Fig. 1. Overall framework based on the gain-field modulation of parietal neurons for coordinate transform and multimodal integration; adapted from Pouget et al. [6]. Parietal neurons translate and coordinate the stimuli information from the visual, the auditory and the proprioceptive signals in eye-, head-, body-reference frame, by varying their gain levels.

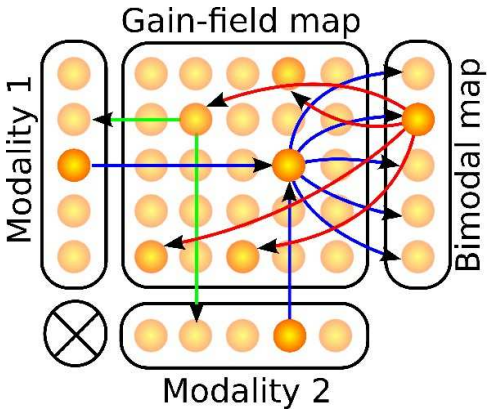


Fig. 2. Reentrant mechanism. The unimodal neurons fed univocal sensory signals to the gain-field neurons and to the downward neurons, and receive back the multimodal response; adapted from [22], [13] and [23].

actions. Despite their respective advantages, both lack the physical embodiment in real robots who face the curse of complexity across the different modalities during motion and temporal integration.

In this paper, we propose to combine these two mechanisms for modeling within robots the development of multimodal integration and spatial cognition in neonates [4], [16]. For this purpose, we use a robot-head with a unique eye and two bionic ears to model visual, audio and proprioceptive integration for objects spatial localization and coordinate transformation. This work pursues several models of the parieto-motor system on which we studied the contributions of motor and spatial development to social cognition [17]. In previous researches, we used spiking neural networks for the detection and learning of synchrony based on the mechanism of spike timing-dependent plasticity [18], either in robotic experiments or in computer simulations [19], [20]. Here, we investigate similar principles using this time the rank-order coding mechanism (ROC) exploited for its rapid spike-based processing [21]. We enlarge its use to the gain-field effect across different modalities.

Our robot head locates visual and audio stimuli relative to their respective reference frame, even during motion. The neurons replicate the gain-field effect of parietal neurons for different spatial locations from audio, visual and proprioceptive pairings. Using a re-entrant mechanism, the processed information is then fed back to the unisensory maps. It follows that the assembled audio-visual signal can then estimate back the position of a stimulus in each modality. We can observe then phenomena such as multimodal enhancement of spatial perception, which can serve for audio-visual speech perception; i.e., correlations between dynamical face and acoustic cues.

II. Architecture and Neural Mechanisms

In this section, we present the neural architecture and the mechanisms that govern the dynamics of the neurons, gain-field modulation and reinforcement learning. We describe first the bio-inspired mechanism of rank-order coding from which we derive the gain-modulated activity of parietal neurons.

A. Rank-Order Coding Algorithm

The Rank-Order Coding (ROC) algorithm has been proposed by Thorpe and colleagues as a discrete and faster model of the derivative integrate-and-fire neuron [21]. ROC neurons are sensitive to the sequential order of the incoming signals; that is, its *rank code*. The distance similarity to this code is transformed into an amplitude value. A scalar product between the input's rank code with the synaptic weights furnishes then a distance measure and the activity level of the neuron. More precisely, the ordinal rank code can be obtained by sorting the signals' vector relative to their amplitude levels or to their temporal order in a sequence. If the rank code of the input signal matches perfectly the one of the synaptic weights, then the neuron fully integrates this activity over time and fires. At contrary, if the rank coding of the signal vector does not match properly the ordinal sequence of the synaptic weights, then integration is weak and the neuron discharges proportionally to it.

The neurons' output v is computed by multiplying the rank of the sensory signal vector I, $rank(I)$, by the synaptic weights w; $w \in [0, 1]$. For a vector signal of dimension M

and for a population of N neurons (M afferent synapses), we have:

$$v_{n\in N} = \sum_{m\in M} \frac{1}{rank(I_m)} w_{m,n} \quad (1)$$

The updating rule of the neurons' weights is similar to the winner-takes-all learning algorithm of Kohonen's self-organizing maps [24]. For the best neuron win and for all afferent signals $m \in M$, we have:

$$\begin{cases} w_{m,win} = w_{m,win} + \Delta w_{m,win} \\ \Delta w_{m,win} = \frac{1}{rank(I_m)} - w_{m,win} \end{cases} \quad (2)$$

Since the synaptic weights follow a power-scale density distribution, the ROC neurons are similar to basis functions, a prerequisite for gain-field modulation; see [5], [13] for a justification proof.

B. Gain-Field Modulation

Gain-field neurons receive the activity-dependent information from two neural population by multiplying unit by unit their value to each other, see Fig. 2 (blue lines). The multiplication between afferent sensory signals from the two population codes, N_1 and N_2, generates the signal activity η_n to the n gain-field neurons, $n \in N_1 N_2$:

$$\eta_n = v_{n_1} \times v_{n_2}. \quad (3)$$

The key information here is the specific amplitude relation between the two neurons. Note that this is a little more subtle than Hebb's law or spiking-or-not activity where neurons are selected only when they have both a high value above a certain threshold. Then, downward efferent neurons can learn the neural activity from the gain-field neurons. By doing so, they realize the encoding of a bimodal information based on the two unisensory signals. The computed mutual information is used next to re-estimate the unisensory signals through a reentry processing stage; see Fig. 2 (red lines). The reentry mechanism is as follows. The triggered pre-synaptic gain-field neurons reinforce their links with the post-synaptic downward neurons; their activity is updated in consequence to have $\eta_n = v_{n_1} \times v_{n_2} + v_n$. This reentry mechanism is similar to the one proposed by [22], [25] for mutimodal integration [23], which can serve then for coordinates transform from one reference frame to another; e.g., auditory or tactile information in eye- or head-centered reference frame.

III. Hardware and Experimental Setup

Our head-robot consists of a box-shaped device mounted on a servo-motor, the neck turns on the sagittal plane and a camera, which is fixed on its eye axis, rolls on the horizontal plane. We plug on the device two bionic ears on which microphones are attached on the eardrums, see Fig. 3. Although the whole system has only two degrees of freedom, the sensory-motor information flow that it can generate (with visual and auditory signals) is already complex enough for modeling difficult coordinate transform problems.

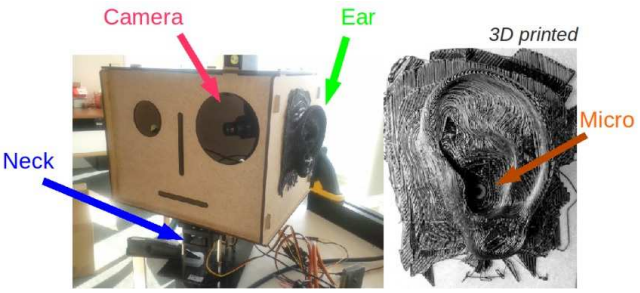


Fig. 3. Our head-robot consists of a head-neck-eye device with ears. The head rotates on its neck and the eye on its axis (left). The 3D-printed bionic ears replicate the shape of human's ears for mimicking human-like spatial localization of audio sources and a similar bandwith filtering of sound's envelope (right).

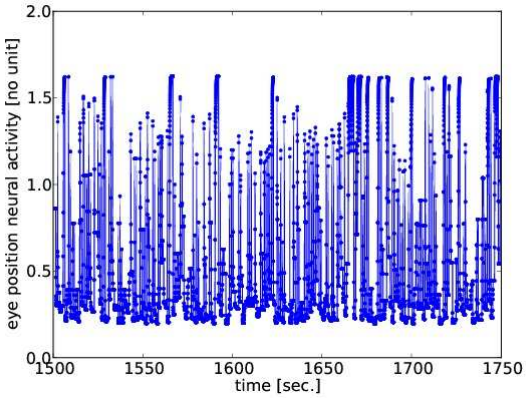


Fig. 4. Neural activity of one eye motor neuron for 250 seconds. Over time, each neuron learns to be selective to one specific motor angle, whose sensitivity is translated as a gain-modulated activity.

The bionic ears have been designed with a 3D-printer based on a 3D model of a human-ear in order to replicate its bio-mechanical characteristics. The microphones can receive an audio signal in the range $[200Hz; 30kHz]$. Moreover, the box-like shape of the head has also a function, it creates a sound *shadow* that eases the discriminating between the left and the right ear. The auditory channel conveys a bank-filter of 40 frequencies selected in the interval $[300Hz; 20kHz]$ following a logarithmic scale to respect the auditory discrimination, toolbox provided by [26].

Considering the visual inputs, we chose an analogic camera to transmit the video signal with a pixels' resolution reduced to $[40 \times 30]$. The motors are moving within the interval $[-60°; +60°]$, and their resolution is discretized to correspond to a 20 bins vector so that each index is associated to one motor angle with a linear scale. Finally, learning is done online in an unsupervised manner with no offline training data.

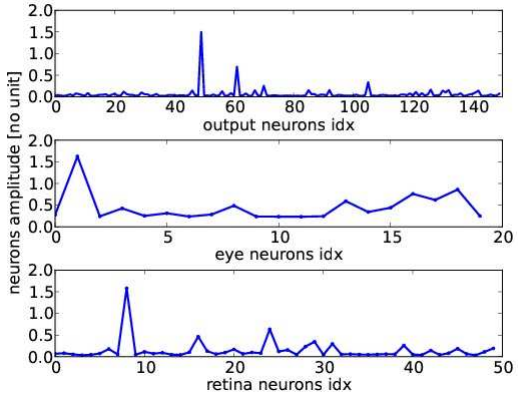


Fig. 5. Snapshot of the vision, eye-motor neural fields and the visuo-motor parietal neurons. A specific parietal neuron (top) is selective to one particular visuo-motor neural pair. Here, the most salient downward parietal neuron (top) is tuned to a motor angle and retina position, resp. the most active motor (middle) and vision neurons (bottom).

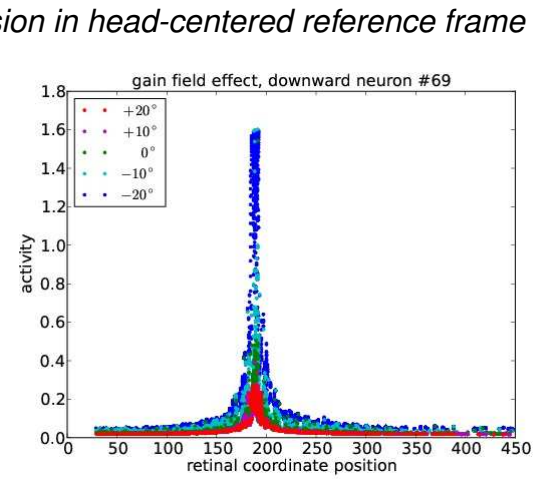


Fig. 6. Gain-field effect relative to visual stimuli localization on the retina for downward neurons #69 (a) and #127 (b). The downward neurons are tuned to certain retinal coordinates, their amplitude is modulated by the motor angles.

IV. Experiments

A. Saccadic Eye-Movement

Our first experiment consists of modeling the visuomotor features of parietal neurons to encode retinal coordinates into a head-centered reference frame using the eye motor signal.

In this setup, we take into account the eye motion only with the visual information, which means that we purposefully ommit the neck and the auditory inputs. The neural population dedicated to the motor-eye signal has respectively 20 neurons (e.g., modality 1 in Fig. 2) and the neural population dedicated to the retina signal has 50 neurons (e.g., modality 2 in Fig. 2) receiving the pixels' activity from the camera. The parietal neurons count therefore $20 \times 50 = 1000$ units (see eq. 3 and the gain-field map in Fig. 2). We add an efferent downward network of 150 units that learns the visuo-motor associations from the afferent parietal neurons activity. Furthermore, each map is initialized with random connections so that all the neurons are at the beginning unspecific to any stimuli.

During the learning stage, at each iteration, the winner neuron of each map (the most salient neuron) sees its synaptic weights updated to shape the receptive field salient to the current entry code. Over time, the neural nets self-organize themselves to map the retina and the eye motor signals. Figure 4 shows the activity of an eye-motor neuron during motion. The neuron's activity describes its selectivity to a specific eye angle, and the firing events occur when the motor response reach a posture close to the neuron's receptive field. At the population level, the neurons responsive to similar visuo-motor signals produce identified activity patterns in the three maps while the cross-product of the visual and motor neural patterns feeds the posterior head-centered neurons, see the snapshots activity in Figures 5.

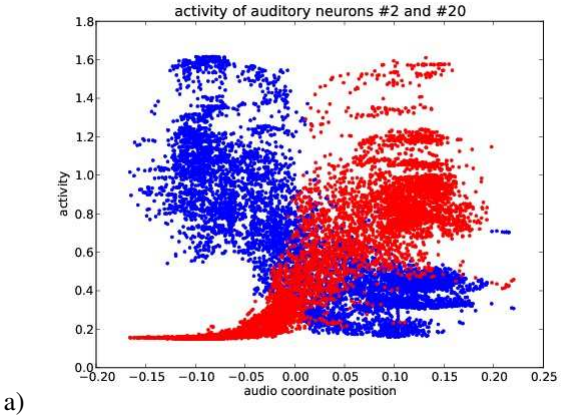
The gain-field effect is observed in Figure 6 for one downward neuron only. The visual receptive fields #69 encodes one retina coordinate in head-centered reference frame so that its position in space is independent of where the eye is fixating (the color index is assigned to one particular motor angle). The neuron is tuned to position pixel 190 and motor angle $-20°$. Its amplitude combines therefore two information at once; a code response similar to ventral intraparietal (VIP) neurons. The linear combination of the downward neurons can be used for tracking behaviours (e.g., for correcting the distance to the eye center) or for translation purpose with other modalities.

B. Auditory Mapping in Head and Body Reference Frames

Although sound information is naturally mapped into a head coordinate system, a consistent proportion of auditory neurons in the parietal cortex exhibits eye-centered and body-centered remapping [27]. That is, the magnitude of the responses for these neurons is modulated respectively by the eye position and the neck movement. For instance, some observations showed that an intended eye movement influences the mapping of the auditory space, and reversely, a perceived sound can influence where to foveate. It is suggested that these behaviours exploit transformation mechanisms such as the one modeled in section IV-A.

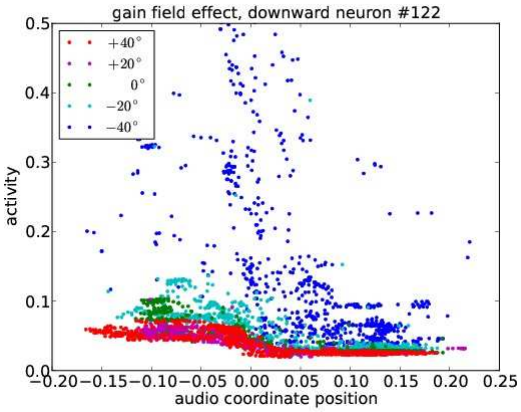
The neural population of the auditory map receives the vector signal of 2×40 frequencies. Then, the sound *shadow* produced by the head permits the rapid self-organizing of the auditory neurons to two distinct receptive fields that discriminate accurately the left and right sides relative to the head horizontal plan, their appropriate reference frame; see the neural activity in Fig. 7 a). The remapping of the auditory signals into a body-centered coordinate system is computed as for the retina/eye-motor transformation in section IV-A: here, the auditory signals and the neck-motor signals modulate a gain-field map that computes the spatial estimation of the sound localization, and this estimation is

Sound in head-centered reference frame



a)

Sound in body-centered reference frame



b)

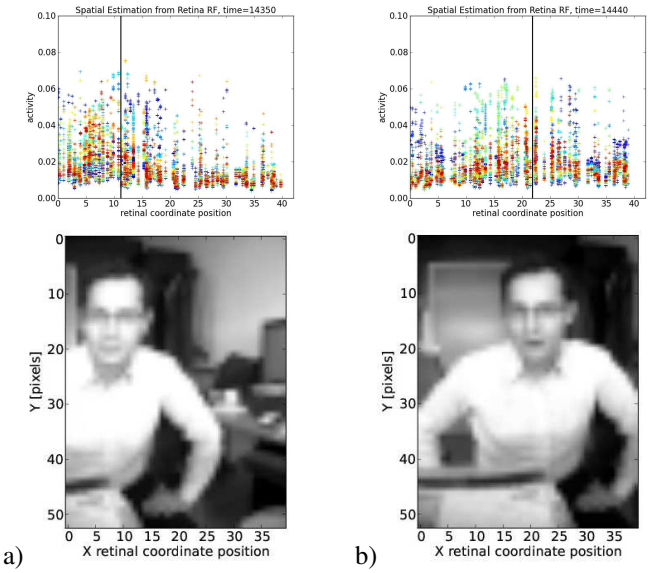
Fig. 7. Sound localization in head- and body-centered reference frames. In a), sounds are naturally mapped into the head-centered reference frame, neurons easily discriminate left and right sides from sound energy intensity. In b), gain-field effect for a downward neuron relative to the neck-motor signals. The auditory stimuli localization from the left and right ears are modulated by the head amplitude signals.

irrespective to the head motion, see the gain-field effect for one downward neuron in Fig. 7 b). As we can observe, the neuron's gain level correlates almost linearly with the sound location. The result is that the referential for sounds is now changed into body-centered coordinates. The neuron is now tuned to a fixed position -40° on the left side of the head. Moreover, in comparison to the head-centered profiles in Fig. 7 a), the neural fields in body-centered coordinates are now enhanced with sharper sound profiles.

C. Audio-Visual Speech Perception

Using the reentry mechanism, mutimodal information can leverage the perceptual processing of unimodal maps in their respective frame of reference to infer spatial location of noisy signals [23]. In our framework, audio-visual information in head-centered reference frame can be transferred back into retinal coordinates within the proper eye-centered reference

Audio-visual speech processing



a) b)

Fig. 8. Super-imposed audio-visual receptive fields in eye-centered coordinates from reentrant signals. The most salient receptive fields are aligned on the top of the camera image; the reddest dots represent the most active retina neurons. In a), the visual information only provides enough information for locating the face correctly. In b), speech vocalization drives the reestimating of the location of the stimulus in space using audio-visual information.

frame for the visual input, in retinal coordinates.

We plot in Figure 8 a) and b) the perceptual processing for audio-visual temporal coherence in retinal coordinates, for facial gestures and acoustic cues. The tuning across the modalities can serve then for locating salient onset and offset signals such as for speech processing in different sensory modalities. In comparison to visuomotor only receptive fields in Fig. 6, and audio only neural fields in Fig. 7, audio-visual mapping takes a position in-between, mixing both information: the center and the variance of the tuning curves.

In a), multimodal information is exploited to estimate the position of the unitary visual stimulus. Here, we super-impose the audiovisual receptive fields in the same abscisse coordinates of the image with a color set 'jet', from the most salient one in red to lowest in blue. The black line indicates the visual stimulus. We observe that the reddest neural fields, whose bubbles are centered on retinal location $X = 10$ in a), are well aligned with the person's face location. In b), during the person's vocalization in a different location relative to the robot's head, the later receives audiovisual stimuli, which are combined to produce a spatial decision. In this situation, the probability distribution of the neural fields are grossly centered on the person.The perceptual system combines each modality in a common shared space to estimate the speech information. By doing so, it shifts the attentional focus to multimodal stimuli, which are understood to be part of the same entity.

We perform statistical analysis from ten minutes data;

10.000 samples. The visual and audio-visual receptive fields overlap for mostly 80% of the time when a visual stimulus is seen in front of the eye field. Thus, we can strictly measure the performance of the system only when the two neural maps differ their estimations. When there is a conflict between the two maps to locate audio-visual stimuli, we observe that audio-visual receptive fields perform two third time better than the visual population, the ratio is 68.75% versus 31.25%. Having audio-visual receptive fields, the performance level of correct location is globally increased by 10% in comparison to the sole retina system with visual receptive fields.

V. Discussions

Seeing engages all our senses. Perceiving one object in space requires to compute its distance relative to our gaze, to our head, or to our hand from multiple sensory types that are not in the same reference frame. Neurobiological observations locate the superior colliculus and the parietal area as the two brain regions where this unified conception of the world could be built [2], [6].

Accordingly, it is interesting to note that these two regions exploit both the same neural mechanism of gain-field modulation [7], [10]. In gain fields, various kinds of combinations of multimodal sensors are represented. In order to organize a reference frame, it is theoretically necessary to bind all possible singleton from all modalities by a multiplicative function. However, only certain combinations of retina and motor are learned, but they are sufficient enough to map the most pertinent sensorimotor experiences.

The gain-field effect is hypothesized to serve for translating the unisensory signals to whichever reference frame, possibly making each modality aligned to another. Thus, the gain-field modulatory mechanism may give some hints on the organizational principle for optimal sensory arrangement; e.g., for compensating the relative motion from the body posture to targets [11]. For instance, it may serve for the construction of the peripersonal space as it is found for the VIP mirror neurons, which integrate many modalities. Although its implication to infant social tasks has not been proposed yet, it may furnish a framework for a coordinate transform mechanism to retranscribe one's body posture to someone else postural configuration, which is a possible link to imitation [28].

Our robotic experiments are preliminary results and we propose to search for more robust solutions within our framework. Also, we will investigate its impact in more complex robot tasks using this time a robot torso with an arm.

Acknowledgment

The authors are grateful to ANR project INTERACT, to Bruno Gas team at ISIR lab. UPMC for the sound drivers, to the FacLab of Cergy-Pontoise University for the laser cut and the 3D printer, to BVS company for the microphones.

References

[1] B. Stein, B. Magalhães Castro, and L. Kruger, "Superior colliculus: Visuotopic-somatotopic overlap," *Science*, vol. 189, pp. 224–226, 1975.

[2] B. Stein and M. Meredith, *The Merging of the Senses*. A Bradford Book, cambridge, MA, 1993.

[3] J. Groh and D. Pai, *Looking at sounds: neural mechanisms in the primate brain*. In, Primate Neuroethology. A. Ghazanfar and M. Platt, eds. Oxford University Press, 2010.

[4] P. A. Neil, C. Chee-Ruiter, C. Scheier, D. J. Lewkowicz, and S. Shimojo, "Development of multisensory spatial integration and perception in humans," *Developmental Science*, vol. 9, no. 5, pp. 454–464, 2006.

[5] A. Pouget and L. Snyder, "Spatial transformations in the parietal cortex using basis functions," *J. of Cog. Neuro.*, vol. 3, pp. 1192–1198, 1997.

[6] ——, "Computational approaches to sensorimotor transformations," *Nature Neuroscience*, vol. 3, pp. 1192–1198, 2000.

[7] E. Salinas and P. Thier, "Gain modulation: A major computational principle of the central nervous system," *Neuron*, vol. 27, pp. 15–21, 2000.

[8] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. Mahwah, NJ: Lawrence Erlbaum Associates, 1949.

[9] R. Andersen and V. Mountcastle, "The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex," *J. Neuroscience*, vol. 3, pp. 532–548, 1983.

[10] E. Salinas and T. J. Sejnowski, "Gain modulation in the central nervous system: Where behavior, neurophysiology and computation meet," *The Neuroscientist*, vol. 7, pp. 430–440, 2001.

[11] S. Chang, C. Papadimitriou, and L. Snyder, "Using a compound gain field to compute a reach plan," *Neuron*, no. 64, pp. 744–755, 2009.

[12] R. Andersen and H. Cui, "Intention, action planning, and decision making in parietal-frontal circuits," *Neuron*, vol. 63, p. 568583, 2009.

[13] S. Deneve, A. Pouget, and J. Duhamel, "A computational perspective on the neural basis of multisensory spatial representations," *Nature Rev. Neurosciences*, vol. 98, pp. 741–747, 2002.

[14] S. Fuke, M. Ogino, and M. Asada, "Acquisition of the head-centered peri-personal spatial representation found in vip neuron," *IEEE Trans. on Aut. Ment. Dev.*, vol. 1, pp. 131–140, 2009.

[15] E. Chinellato, M. Antonelli, B. Grzyb, and A. del Pobil, "Implicit sensorimotor mapping of the peripersonal space by gazing and reaching," *IEEE Trans. on Aut. Ment. Dev.*, vol. 7, no. 3, pp. 43–53, 2011.

[16] D. Lewkowicz, "Infant perception of audio-visual speech synchrony," *Developmental Psychology*, vol. 46, no. 1, pp. 66–77, 2010.

[17] A. Pitti, H. Mori, Y. Yamada, and Y. Kuniyoshi, "A model of spatial development from parieto-hippocampal learning of body-place associations," *10th Inter. Conf. on Epigen. Rob.*, pp. 89–96, 2010.

[18] L. Abbott and S. Nelson, "Synaptic plasticity: taming the beast," *Nature neuroscience*, vol. 3, pp. 1178–1182, 2000.

[19] A. Pitti, H. Alirezaei, and Y. Kuniyoshi, "Cross-modal and scale-free action representations through enaction," *Neural Networks*, vol. 22, no. 2, pp. 144–154, 2009.

[20] A. Pitti, H. Mori, S. Kouzuma, and Y. Kuniyoshi, "Contingency perception and agency measure in visuo-motor spiking neural networks," *IEEE Trans. on Aut. Ment. Dev.*, vol. 1, no. 1, p. 8697, 2009.

[21] R. Van Rullen and S. Thorpe, "Surfing a spike wave down the ventral stream," *Vision Research*, vol. 42, pp. 2593–2615, 2002.

[22] P. Roelfsema and A. van Ooyen, "Attention-gated reinforcement learning of internal representations for classification," *Neural Computation*, vol. 17, pp. 2176–2214, 2005.

[23] J. Driver and C. Spence, "Multisensory perception: Beyond modularity and convergence," *Curr. Bio.*, vol. 10, no. 20, pp. R731–R735, 2000.

[24] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

[25] S. Deneve and A. Pouget, "Bayesian multisensory integration and cross-modal spatial links," *J. of Phys.–Paris*, vol. 98, pp. 249–258, 2004.

[26] M. Bernard, S. N'Guyen, P. Pirim, B. Gas, and J.-A. Meyer, "Phonotaxis behavior in the artificial rat psikharpax," in *Int. Symp. on Rob. and Int. Sensors, IRIS2010*, Nagoya, Japan, 2010, pp. 118–122.

[27] B. Stricanne, R. Andersen, and P. Mazzon, "Eye-centered, head-centered, and intermediate coding of remembered sound locations in area lip," *J. Neurophysiol.*, vol. 76, pp. 2071–2076, 1996.

[28] A. Meltzoff, "like me: a foundation for social cognition," *Developmental Science*, vol. 10, no. 1, pp. 126–134, 2007.