# Distinct Mechanisms for Multimodal Integration and Unimodal Representation in Spatial Development

Alexandre Pitti, Arnaud Blanchard, Matthieu Cardinaux, Philippe Gaussier

# Distinct Mechanisms for Multimodal Integration and Unimodal Representation in Spatial Development

Alexandre Pitti, Arnaud Blanchard, Matthieu Cardinaux and Philippe Gaussier

***Abstract*—In this paper, we attempt to reconciliate two views of spatial development based on two mechanisms of statistical learning and of sensory alignment. Conflicting results in developmental psychology attribute either a developmental period to spatial cognition (Piaget). Besides, these results conflict with other researches in which infants do demonstrate good coordination and coherence across modalities (Gibsonian), even from restricted pre-natal experiences [1], [2].**

**In order to study both views, we present at first a simple model based on conditional learning which integrates visual and auditory modalities although it has some limitation regarding the number of degrees of freedom. In second, we propose then to use a sensory alignment mechanism, which allows the system to learn invariances in the world. In experiments with a robot head, we show the advantages of each strategy. We then discuss about the future possibilities of merging both models and their implications.**

***Index Terms*—robotics, mutimodalities, learning, gain-field, audio, vision**

## I. Introduction

Perceiving objects in space is one of the 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), and sometimes encoded in different reference frames (e.g., eye-centered, head-centered or hand-centered). This curse of dimensionality corresponds to the so-called binding ability across modalities for which there is still debates on its underlying neural mechanisms and its associated computational models.

Piaget emphasizes particularly this problem because it is until the second half of the first year that infants appear to fully overcome features binding as they begin to apprehend sequences of actions as unified events: at this period, they start to recognize that the changes they perform in the world have causal consequences that permit them to think about physical events [3]. Many researchers recognize that statistical learning mechanisms play a key role in detecting the regularities that are necessary to build the causal links across the modalities during perceptual experiences [4], [5], [6]. Some recent papers even argue for a later maturation in middle childhood [7].

Besides, these results conflit with other researches in which neonates do demonstrate coordination and coherence across modalities, even from restricted pre-natal experiences [1], [2]. Post-natal experiments done by Gibson, Spelke and Meltzoff (among other researchers) on intermodal matching argue for such a view [8], [9], [10]. It advocates for the existence of efficient structuring mechanisms for organizing coherently the sensory signals from each others. Accordingly, auditory, tactile or visual modalities would be perceived inter-dependent and unified –, such as a unimodal map [11],– and not just serially connected through statistical and adaptive mechanisms. In this paper, we attempt to reconciliate both views and show that the two mechanisms, statistical learning and sensory alignment, may co-exist and work in concert toward cognitive development. For instance, statistical learning can explain the situations where infant links gradually a modality to the other, while the neural mechanisms for coherent sensory alignment would explain why infants recognize from birth the dependency across sensory signals (e.g., for recognizing his mother's voice and face [12]).

In neuroscience, this debate on the organizational principles of intermodality pervades also. On the one hand, collicular neurons, parietal and motor neurons (the mirror neuron system) demonstrate that sensory-motor integration is achieved at a "lower" level at which motor neurons integrate not only visual but also auditory information about the location of objects within peripersonal space [13]. Gallese argues that no common shared space is necessary to unite the modalities, he proposes rather that the associations are done through reinforcement learning which accurately merges the long-range synaptic links between each modality. For instance, the raw calibration of sensory-motor contingencies during motor babbling can permit the infant to build statistical correlations between his sight, his hearing and his touch [14]. Other experiments show that infants look almost exclusively at their hands only when they are intentionally moved [15] and that targeting the mouth during speech perception appears only during the fourth month [16]. Nevertheless, such a view does not explain *(1)* how reinforcement learning organizes efficiently the sensory signals and *(2)* how it can help for the construction of a bodily frame of reference for action [15].

On the other hand, neural circuits such as the superior colliculus and the parietal cortex show properties of sensory registration on a univocal reference frame, centered on the head, the eye [17], [18] or on the body [19]. Interestingly, these two maps rely on a specific mechanism, known as the gain-field modulatory effect [20], [21], [22], that binds the modalities from each other and translates each coordinate space without any necessary adaptation period, which is a feature that is potentially important to maintain a mapping of sensory coordinates of objects into motor coordinates [21]. For instance, this mechanism may permit infants to locate objects in a cartesian-like absolute space, seemingly free from the body posture and motion. Such mechanism might explain why infants process faster audiovisual signals than unimodal signals [5].

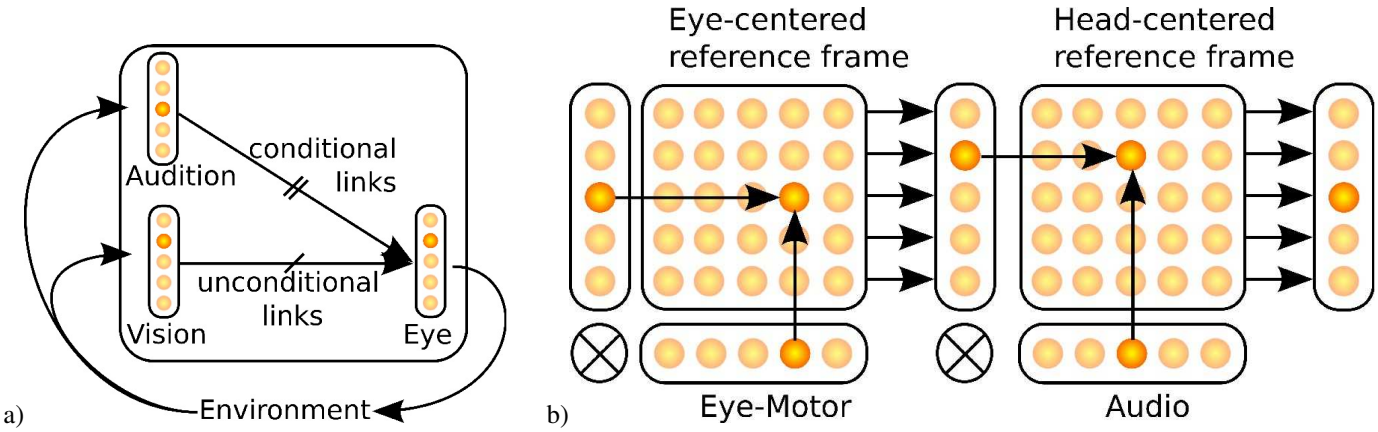


Fig. 1. (a) Overall framework for multimodal integration via conditional learning. Vision has conditional links to the eye and audition has unconditional links. (b) Overall framework for a unimodal architecture; adapted from Pouget et al. [23]. Gain-field modulation in parietal neurons permits the coordinate transform from one modality to the other in each reference frame by varying their gain level, because their are in registry. One result is the observation of a multisensory enhancement of spatial location by re-arranging the reference frames correctly (e.g., head-centered reference frame).

The development of these two neural mechanisms would explain why two different spatial mechanisms co-evolve during infancy [24]. The first mechanism, the timing integration in sensory-motor circuits, means the detection of temporal events like synchrony or rhythmicity, see Fig. 1 a). It leads the perceptual enhancement or discrepancy in attentional tasks by reinforcing statistically the links of contingent neurons, as emphasized in Hebb's law [25]. The second mechanism, the gain-field modulation, describes the phenomenon where the motor and the sensory signals are in registry from each other: each visuo-motor pair mutually influences –,or modulates,– the amplitude activity of their corresponding afferent neurons, see Fig. 1 b). This feature observed in posterior parietal neurons (PNNs) permits to link one eye position with a retinal location, independently to any specific reference frame. Differently said, these neurons can be used to encode stimulus location in more than one reference frame using "gain fields" [21].

In this paper, we compare these two neural mechanisms for modeling within robots the parallel development of multimodal integration and spatial cognition in neonates [12], [26], [6]. We use for this a robot-head to investigate visual, audio and proprioceptive integration. For the mechanism of reinforcement learning, we show how simple conditional rules across audio and visual modalities permit to make some expectations about stimuli's location in each modality and to make some goal-directed behaviors (e.g., to direct gaze). We emphasize here the learning of links which binds statistically the modalities from each others with goal-directedness, even though there is not a strict inter-dependence across the modalities, which is not necessary here. Besides, for our second mechanism, no learning phase is provided for binding, we show how the strict inter-dependence across the modalities permits to remap spatial features from an initial reference frame to the desired one (e.g., vision in head-centered reference frame). Changing reference frame permits to combine modalities from each others in a common space and to infer more accurately one intermediate estimation; i.e., multimodal enhacement [27].

Although these two mechanisms appear to contradict, they encompass the existing conflicting results encountered in developmental psychology and in developmental cognitive neuroscience between Piagetian and Gibsonian. We discuss how they may co-evolve separately during the first year to be the starting blocks for spatial development. Furthermore, we speculate also about the neural structures they may correspond to. We present the neural mechanisms that govern the dynamics of the two neural architectures for sensory-motor learning and modality alignment; i.e., reinforcement learning and gain-field modulation. We apply these mechanisms on a robot that learns to track objects or persons using sound and vision. The object will be noisy and shaking (typically a rattle) in order to activate simultaneously auditory (sound energy) and visual ( saliency, movement detection ) modalities (see Fig. 2).

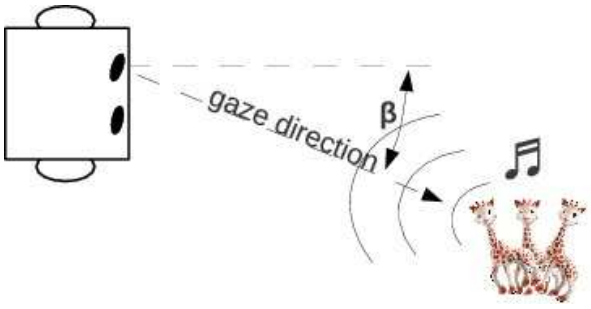


Fig. 2. Setup: the robot has to orient its gaze toward the stimuli using either vision or audition.

In section II we will test on a robotic head the multimodal integration based on conditioning and in section III we will test the effect of the unimodal map based on gain field modulation.

## II. Multimodal Integration based on conditioning

In this part we propose an architecture allowing to associate the audio-modality with the visual-modality.

### A. Architecture and Neural Mechanisms

In the PerAc architecture –, acronym for Perception-Action [28],– the idea is to use an innate behavior (here the visual tracking of moving objects) as an unconditional input for an associative learning rule, see Fig. 1 a). We can make the robot progressively use sound to track moving and noisy objects. The auditive modality can progressively improve the tracking behavior as newborns learn to do it [5].

In order to implement this behavior, we use the difference of sound energy in the two ears as the auditive cue. We implement a visual tracking reflex based on movement detection (difference of luminosity between two successive images). The robot makes saccadic eye movement each second (time for the camera to stabilize) depending on the position of the movement in the retina. At the same time it analyses the difference of sound energy in each ear and learns to predict the eye orientation. The learning rule is based on Widrow and Hoff algorithm (see eq.2).

$$E = \sum_{i \in \text{conditional links}} S_i.W_i \quad (1)$$

$$\Delta W_i = \epsilon.(E - S_i.W_i) \quad (2)$$

$E$ is a neuron coding for a specific eye orientation, $W_i$ the weight between this neuron and the input neurons $S_i$ coding for sound difference. $\epsilon$ is the learning rate.

After learning, the vision module progressively predicts the eye position. The final movement of the eye is a neural field summing the request of the vision and the audition. At the beginning the prediction of the audition will be very low and therefore the vision will be dominant but gradually audition will also contribute and may replace vision when the target is out of the visual field or if the eyes are closed.

### B. 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 fixated 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 and multimodal integration. 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. For instance, the asymmetrical and curvy shape of the human ear allows to filter the proper bandwidth frequencies of the auditory signals for interacting socially with people (e.g., to detect human voices) and physically with objects (e.g., to locate the position of noisy objects).

More precisely, the microphones can receive an audio signal in the range $[200Hz; 30kHz]$ but we measured that the shape of the ear amplifies qualitatively the signal in the interval $[10kHz; 14kHz]$ and shifts its wave envelope in the higher frequencies (for technical details see [29]). This morphological computation enhances the bandwidth filtering, spreads the signal with more frequencies in the higher tone which is useful for the afferent signal processing of the audio sources, like sound separation and pattern discrimination. Moreover, the box-like shape of the head 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. It is based on a cochlear model [30].

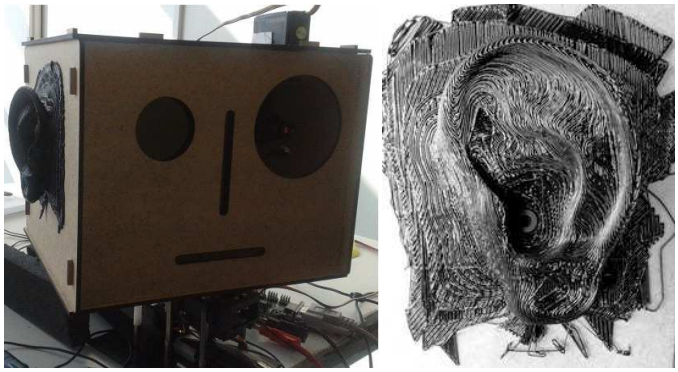

Fig. 3. Our head-robot consists of a head-neck-eye device with ears. The head rotates on its neck and the eye on its axis (left). The 3D-printed bionic ears replicate the shape of human's ears for mimicking human-like spatial localization of audio sources and a similar bandwidth filtering of sound's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

For the visuo-proprioceptive inputs, we use an analogic camera to transmit the video signal with a pixels' resolution reduced to $[40 \times 30]$. Besides, 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. We note that, in order to smooth a little the motor vector, we add a surrounding bubble activity centered around the current motor index, following a gaussian neighbouring law.

### C. Experiment

In this experiment we control the position of a noisy and moving object (a speaker that we shake) at the front of the robotic head. The eye makes saccadic movements each second in order to visually track the object ( orienting itself to the maximum of differences between two images ). We code the neck to oscillate (with saccades also in order to avoid permanent movements on the camera). The eye has to correct this oscillation in order to follow the target. During this time the auditive system is learning to associate the difference of sound energy and the position of the eye ( presented in Fig 1.a ) . After a while if the object is not shaking any more, or if the camera is in the dark, the tracking is still possible using the auditive signal even though the link between sound and position has never been encoded.

In Fig 4 we present the result of an experiment where we see that at the beginning the prediction of the real object in solid line based on sound in dash-dot line is random. However after learning to associate the sound and the position of the eye (in dash line), the prediction is getting better. After 70 seconds when we stop the camera, the tracking of the noisy object is still possible. The precision is not perfect because the time of learning is very short and because the sound is not as precise as vision.

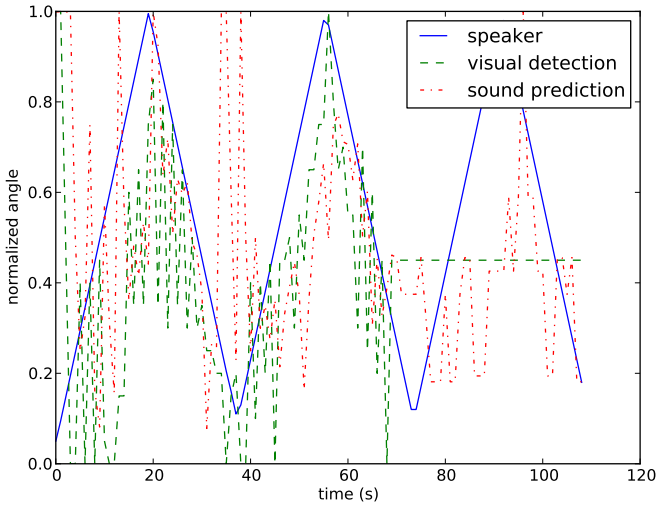


Fig. 4. When we move a noisy object at the front of the robot, the orientation of the gaze is at the beginning only based on vision. After learning, audition is also able to predict the position of the object. When after 70 seconds we stop the visual system, the robot is still able to orient the attention to the object.

This system is very efficient and simple but would not work when we need to consider a combination of degrees of freedom (i.e. body, neck plus eye), but the unimodal approach would permit it.

## III. Unimodal Map through Gain-Field Modulation

The formation of an associative map, as it is seen in parietal and collicular circuits, is created in our framework with the mechanism of gain-field modulation.

The mechanism of gain-field modulation is as follows. Efferent parietal neurons receive the activity-dependent information of each neural population by multiplying unit by unit their value to each other, which is the main difference with the previous circuit. The multiplication between afferent sensory signals from the two population codes, $N_1$ and $N_2$, generates the activity of $n$ parietal neurons $\eta_n$, $n \in N_1 N_2$:

$$\eta_n = v_{n_1} \times v_{n_2}. \tag{3}$$

By doing so, the synthetic parietal neurons realize the encoding of a mutual information between two modalities in a sort of central "associative map", which is therefore supramodal. This neural population can serve then for coordinate transform from one reference frame to another; e.g., vision and audio, the eye and the head, the ears and the head.

### A. Hardware and Experimental setup

The hardware is the same robotic head than the one described in section II-B. The visual detection of the target is done using its saliency in the scene in order to avoid the need of moving all the time the target (a speaker).

### B. Experiments

The spatial incoherence between visual stimuli (in eye-centered reference frame) and the auditory sources (in head-centered reference frame) has to be resolved either with adaptation mechanisms (see section II) or with coordinate transform mechanisms (in this section). For instance, to transform retinal coordinates into head-centered reference frame, parietal neurons require to combine first the eye motor signal with retina signals. It is only then that integration with auditory sources is possible in the same spatial map; see Fig. 1. The neural population dedicated to the eye-motor signal has respectively 20 units to categorize the motor space and the neural population dedicated to the retina signal has 50 neurons with 1200 synapses receiving the pixels' activity from the camera. The parietal neurons count therefore $20 \times 50 = 1000$ units (see eq. 3). The downward network possesses 150 units and receives the activity of the parietal neurons. Each unit learns a particular visuo-motor association from the parietal neurons. That is, at the neuron's level, each unit shows selectivity to their specific input signals: motor neurons to motor angles, retina neurons to pixel's saliency and downward neurons to particular retina-motor pairs.

Here, The gain-field effect is observed in Figure 5 for a downward neuron. Its visual receptive field encode one particular retina coordinates in head-centered reference frame so that its position in space is independent of where the eye is fixating: the color index is assignated to a particular motor angle and for different motor angle, the camera still continues to see the salient stimulus but in peripheral view. Its amplitude furnishes at once information about the two modalities, similar in response to ventra-intraparietal (VIP) neurons [31]. These neurons can be used then for tracking behaviours or for translation purpose with other modalities in same or different reference frames (e.g., skin or auditory signals).

### C. Multisensory Enhancement of Spatial Location

Multimodal integration is easier to achieve if the modalities are represented in the same format. For instance, vision signals in head-centered reference frame are easier to bind with auditory information, also in head coordinates. The mechanism of gain-field modulation is re-employed here for multiplying the two neural fields. And since they are basis functions like gaussian or sigmoid curves, their combination forms also a basis function [27]. We want to have from their merging a gain

### *Vision in head-centered reference frame*

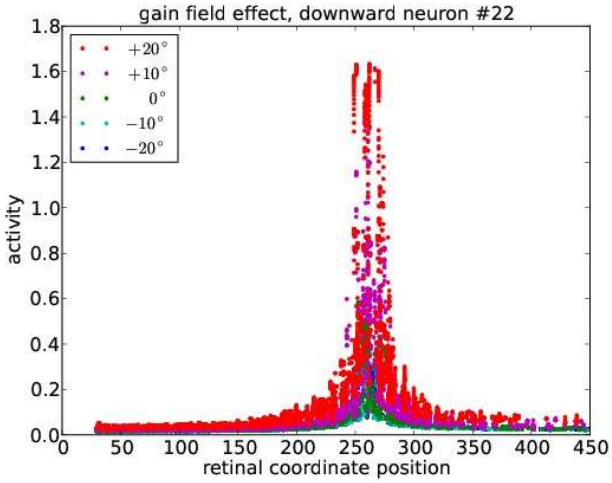


Fig. 5. Gain-field effect in a downward neuron relative to visual stimuli localization on the retina for different motor angles. The activity of the neuron is tuned to certain retinal coordinates and its amplitude is modulated by the motor angles.

### *Sound in head-centered reference frame*

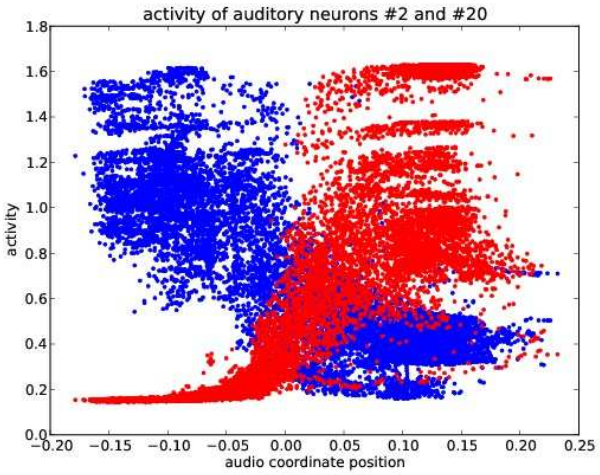


Fig. 6. Activity of two auditory neurons from the left and right ears and spatial estimation of one auditory source in head-centered reference frame. The neurons discriminates easily the symmetry of the head in the transversal plane as they are tuned to the left and right locations.

in spatial estimation, a phenomenon known has multisensory enhancement [6].

The neural population of the auditory map receives the vector signal of $2 \times 40$ frequencies. Due to the strong bias of the head morphology, its neurons rapidly self-organize to two distinct receptive fields and approximate smoothly two sigmoids that discriminate accurately left and right side relative to the head transversal plan; see the neural activity in Fig. 6.

At the next processing stage, the parietal neurons receive the joint activity of auditory neurons and of head-centered vision neurons. They are therefore highly influenced from where the robot *sees* and from where the robot *hears* the audio sources. Their spatial estimation of a simulus source is processed by the multiplication of the gaussian-like visual receptive fields and the sigmoid-like auditory neural fields. Three cases occur depending on where audio-visual stimuli are estimated in space, see Fig. 7. First, when the spatial location of a bimodal stimulus is estimated at the same place, we observe an amplitude enhancement of the parietal neurons that is obviously higher than the two feeding signals. Reversely, a large dissonance between the two modalities disrupts the synchronization matching of the parietal neurons whose amplitude are strongly damped (perceptual discrepancy). Finally, an interesting situation happens when the two estimations are nearby from each other since the predicated location is averaged from both modalities. Ground truth investigations demonstrate a gain on spatial estimation of audio-vision stimuli over unisensory signals of ten percents.

### *Multisensory estimation*

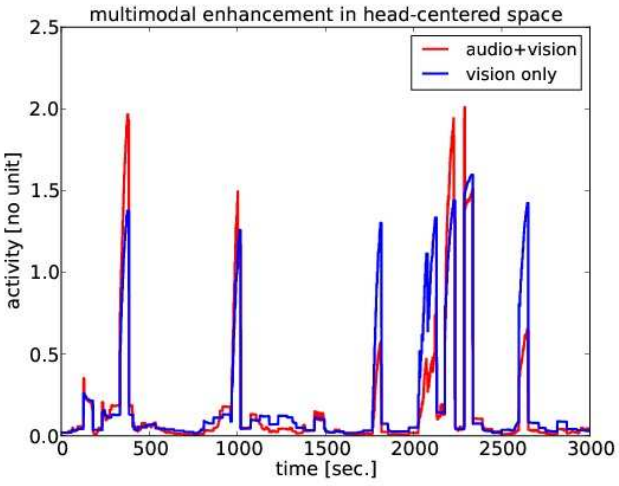


Fig. 7. Estimation of a spatial location from the activity of a vision neuron in head-centered reference frame (blue) and a multimodal neuron (red). Audition influences the prediction of a visual stimulus in space. Coherent audio and visual predictions enhance the estimation, while the modalities' mismatching leads to discrepancy.

## IV. Discussions

We presented two mechanisms for representing space using opposite strategies. The first mechanism is based on reinforcement learning for which contingency detection across multiple modalities serves to correlate them from each other. It is well documented now that adaptation to audio-visual stimuli occurs very early in infancy and serves for motor development.

Learning to predict repeated occurrences of certain stimuli and specific spatial locations (conditional learning) is important for assimilating spatial and motor habits. We hypothesize that the subcortical maps of the cerebellum, the striatum and the basal ganglia are involved in developing these skills based on reinforcement learning and raw sensory-motor anticipation. Besides, it is not clear how efficient this strategy would be when the dimensionality augments. Computing accurately coordinates transform (e.g., for mental rotation) may require other mechanisms for which we have the sensory modalities in registry. The inter-dependencies, bred by the gain-modulation mechanism, binds univocally the modalities to perform dual coordinate transform. For instance, it has been found receptive fields in the parietal area combining eye-hand distance neurons for planning [32]. The compound coordinate transform is no more centered in one location but changes dynamically with the two variables system.

Neonates do not seem capable to perform these complex operations as they appear to learn gradually their whole body representation in multiple reference frames. Studies in mammals neonates by Barry Stein show that the superior colliculus has its modalities encoded in eye- and head-centered coordinates. However, babies seem capable to understand very rapidly that objects and persons are coordinated systems: for instance, they are sensitive to biological motion and to facial configurations even at an early stage. They do not seem to learn statistically the relationships between each variable through a learning stage but rather to understand the invariance of the coordinated system seen from different angles. Similarly, such mechanism may permit to detect also the invariances in facial configuration; the eyes, nose and mouth invariant locations, as well as the mouth and voice invariant reference frame.

## ACKNOWLEDGMENTS

The authors are grateful to the FacLab from the University of Cergy-Pontoise for the laser cut and the 3D printer, to Olivier Gendrin (FacLab manager) for the valuable help he gave to make the head and to BVS company for the cochlear model. This work is part of the project INTERACT (ANR09-COORD014), program CONTINT. (http://interact.ensea.fr)

## REFERENCES

[1] H. Prechtl, "Prenatal and early postnatal development of human motor behaviour," *Handbook of brain and behaviour in human development. Kalverboer AF, Gramsbergen A, editors. Amsterdam: Kluver*, pp. 415–427, 2001.

[2] J. Lepage and H. Théoret, "The mirror neurons system: grasping others' action from birth?" *Developmental Science*, vol. 10, pp. 513–523, 2007.

[3] R. Baillargeon, J. Li, Y. Gertner, and D. Wu, "How do infants reason about physical events?" *in The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition (ed U. Goswami), Wiley-Blackwell, Oxford, UK*, pp. 11–48, 2010.

[4] L. Smith and A. Sheya, "Is cognition enough to explain cognitive development?" *Topics in Cognitive Science*, pp. 1–11, 2010.

[5] 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.

[6] B. Stein, T. Perrault Jr., T. Stanford, and B. Rowland, "Postnatal experiences influence how the brain integrates information from different senses," *Frontiers in Integrative Neuroscience*, vol. 30, no. 14, pp. 4904–4913, 2010.

[7] M. Ernst, "Multisensory integration: A late bloomer," *Current Biology*, vol. 18, no. 12, pp. R519–R521, 2008.

[8] J. Gibson, *The senses considered as perceptual systems*. Boston: Houston Mifflin, 1966.

[9] E. Spelke and C. Owsley, "Intermodal exploration of knowledge in infancy," *Infant Behavior and Development*, vol. 2, pp. 13–27, 1979.

[10] A. Meltzoff and R. Borton, "Intermodal matching by human neoinates," *Nature*, vol. 282, pp. 403–404, 1979.

[11] A. Meltzoff, "Explaining facial imitation: A theoretical model," *Early Development and Parenting*, vol. 6, pp. 179–192, 1997.

[12] J. Hammond, "Hearing and response in the newborn," *Developmental Medecine and Child Neurology*, vol. 12, no. 1, pp. 3–5, 1970.

[13] V. Gallese and G. Lakoff, "The brains concepts: The role of the sensory-motor system in reason and language," *Cognitive Neuropsychology*, vol. 22, pp. 455–479, 2005.

[14] P. Rochat, *The Infant's World*. Harvard Press, 2001.

[15] A. van der Meer, F. der Weel, and D. Lee, "The functional significance of arm movements in neonates," *Science*, vol. 267, pp. 693–695, 1995.

[16] D. Lewkowicz and A. Hansen-Tift, "Infants deploy selective attention to the mouth of a talking face when learning speech," *Proceedings of the National Academy of Sciences, Early Edition*, vol. 10.1073/pnas.1114783109, pp. 1–6, 2003.

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

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

[19] M. Graziano, X. Hu, and C. Gross, "Visuo-spatial properties of ventral premotor cortex," *J. Neurophysiol.*, vol. 77, p. 226892, 1997.

[20] A. Van Opstal, K. Hepp, Y. Suzuki, and V. Henn, "Influence of eye position on activity in monkey superior colliculus," *Neurophysology*, vol. 74, pp. 1593–1610, 1995.

[21] A. Pouget and L. Snyder, "Spatial transformations in the parietal cortex using basis functions," *Journal of Cognitive Neuroscience*, vol. 3, pp. 1192–1198, 1997.

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

[23] A. Pouget and L. Snyder, "Computational approaches to sensorimotor transformations," *Nature Neuroscience*, vol. 3, pp. 1192–1198, 2000.

[24] A. Bremner, N. Holmes, and C. Spence, "Infants lost in (peripersonal) space?" *Trends in Cognitive Sciences*, vol. 12, no. 8, pp. 298–305, 2008.

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

[26] L. Yu, B. Rowland, and B. Stein, "Initiating the development of multisensory integration by manipulating sensory experience," *Journal of Neuroscience*, vol. 30, no. 14, pp. 4904–4913, 2010.

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

[28] P. Gaussier and S. Zrehen, "Perac: A neural architecture to control artificial animals," *Robotics and Autonomous Systems*, vol. 16, no. 2-4, pp. 291–320, 1995.

[29] A. Pitti, A. Blanchard, M. Cardinaux, and P. Gaussier, "Gain-field modulation mechanism in multimodal networks for spatial perception," in *International Conference on Humanoid Robots*, Osaka, Japon, 2011, in press.

[30] M. Bernard, S. N'Guyen, P. Pirim, B. Gas, and J.-A. Meyer, "Phonotaxis behavior in the artificial rat psikharpax," in *International Symposium on Robotics and Intelligent Sensors, IRIS2010*, Nagoya, Japon, 2010, pp. 118–122.

[31] C. Colby and J. Duhamel, "Ventral intraparietal area of the macaque: anatomic location and visual response properties," *Journal of Neurophysiology*, vol. 69, no. 3, pp. 902–914, 1993.

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