

# Perception multimodale de l'espace

D. Philipona <sup>\*,†</sup>

david.philipona@polytechnique.org

J.K. O'Regan<sup>†</sup>

http://nivea.psych.univ-paris5.fr

L'étude de la perception passe-t-elle inévitablement par la taxinomie des éléments neuronaux associés aux différentes modalités sensorielles ? La question paraît saugrenue à nombre de neurobiologistes, pourtant elle découle naturellement de certaines expériences et de réflexions fondamentales qui remettent en cause notre compréhension de la perception.

Selon ce qu'on peut appeler *l'explication passive*, le stimulus physique induit une modification de l'état des neurones sensoriels reliés à des aires corticales dont les activations provoquent la perception et la création d'une sensation. Or cette explication souffre de plusieurs insuffisances. D'une part, le rôle de l'action semble essentiel au niveau le plus fondamental de la perception [3, 5, 7], ce qui s'articule difficilement avec le schéma passif. D'autre part, ce schéma ne pourrait fournir à lui seul une explication complète que si les aires corticales matures avaient une structure et une identité prédéterminée. Ceci est toujours sujet de discussion [9, 10], et il est connu depuis longtemps que le développement du système nerveux n'est pas indépendant des stimulations qu'il expérimente [4]. Il ne s'agit donc pas véritablement d'une réponse mais de la substitution d'une question à une autre : si la perception est l'activation de certaines structures nerveuses, comment ces structures émergent-elles de l'adaptation à certains stimuli ? Finalement, l'explication passive est même critiquable en amont de toute expérience : l'argument selon lequel "il est dans la nature de certaines aires corticales de produire certaines sensation" rappelle la physique aristotélicienne ("il est dans la nature des masses de tomber") qui marquait plus la frontière du cadre scientifique de l'époque que l'accès à une vérité ontologique. On peut ainsi se demander pourquoi il semble acceptable d'imaginer que les neurones de V1 à V4 produisent des sensations lorsqu'il serait évident que les neurones rétiniens en sont incapables.

L'une des alternatives proposées à l'explication passive est la *théorie des contingences sensorimotrices* [7]. Cette théorie réintègre l'action et le stimulus physique au cœur de la perception, en posant comme explication des sensations les propriétés de la dépendance sensorimotrice (la relation entre les ordres moteurs et le retour sensoriel) plutôt que les entrées sensorielles seules, résultats d'une acquisition purement réceptive. Cette approche - contrairement à l'explication passive dont l'objectif est l'établissement d'une *classification* des sous-systèmes neuronaux, de leurs connexions, des différents types de neurones, des phénomènes chimiques et moléculaires impliqués dans leur fonctionnement - fournit un cadre de réflexion sur les *principes* de la perception, cadre synthétique et unifié.

Nous présentons ici la situation du modèle mathématique que nous cherchons à développer pour exprimer les idées de la théorie sensorimotrice, ainsi que les résultats obtenus vis-à-vis de la perception de l'espace. Il s'agissait essentiellement de formaliser la notion de loi sensorimotrice et de propriétés intrinsèques de cette loi. Nous nous sommes donc mis "à la place" de la fonction (l'algorithme) implémentée par le système nerveux central, donnant des ordres moteurs à un certain corps, recevant en retour des inputs sensoriels, avec comme objectif de chercher ce qui pouvait être compris de derrière cette interface inconnue

---

\*Sony CSL, 6 rue Amyot, 75005 Paris

†Laboratoire de Psychologie Expérimentale, CNRS, Université René Descartes, 92774 Boulogne-Billancourt Cedex

impliquant un système sensorimoteur inconnu dans un monde inconnu. Supposer l'un des maillons connu reviendrait en effet à accepter en partie l'explication passive, en mettant cette connaissance implicite dans la structure prédéterminée du cerveau. Comme on l'a déjà dit, ceci ne fait que déplacer le problème vers l'explication de cette prédétermination et ne nous semble donc pas satisfaisant. Il ne faut pas pour autant prendre le point de vue que nous épousons comme une prise de position dans le débat nativiste/empiriste sur les connaissances innées des organismes vivants : nous disons simplement que la position nativiste laisse entièrement ouverte la question de l'émergence de ces connaissances innées et est donc au mieux incomplète pour expliquer les mécanismes de la perception.

## 1 Formulation fonctionnelle de la théorie sensorimotrice

Notons  $S$  les inputs sensoriels, et  $M$  les outputs moteurs. Ce sont les seules choses auxquelles l'algorithme a accès. Pour l'exposé de la démonstration, nous noterons  $P$  les configurations du "corps" contrôlé par l'algorithme et  $E$  les configurations de l'environnement.

Nous allons supposer que le corps est contrôlé, via une certaine fonction  $\varphi_a$ , par l'output moteur  $M$  (dans un cadre neurobiologique, on pourrait par exemple imaginer qu'il s'agit d'un vecteur de  $\mathcal{R}^m$  dont la  $i^{me}$  coordonnée serait l'activité du  $i^{me}$  neurone moteur, ou tout autre grandeur(s) caractéristique(s) de son état). Nous allons également supposer que les inputs sensoriels  $S$  (dans un cadre neurobiologique, par exemple un vecteur de  $\mathcal{R}^s$  décrivant les activités des neurones sensoriels), sont déterminées par une certaine fonction  $\varphi_b$  de la configuration du corps et de la configuration de l'environnement :

$$P = \varphi_a(M) \quad \text{et} \quad S = \varphi_b(P, E)$$

Nous noterons  $\varphi(M, E) = \varphi_b(\varphi_a(M), E)$ , et appellerons cette fonction "loi sensorimotrice fonctionnelle", la fonction qui relie, dans ce modèle, les inputs sensoriels aux outputs moteurs et à la configuration de l'environnement. Nous noterons par ailleurs  $\mathcal{S}, \mathcal{M}, \mathcal{P}, \mathcal{E}$  l'ensemble de tous les  $S, M, P, E$ , et nous formulons une hypothèse de régularité sur ces ensembles et sur la fonction sensorimotrice, précisément que  $\mathcal{M}, \mathcal{E}$  et  $\mathcal{S}$  sont (au moins dans un certain voisinage) des variétés et que  $\varphi$  est une application  $\mathcal{C}^\infty$  vis à vis de la structure de ces variétés.

### 1.1 Structure invariante de la loi sensorimotrice

Dans la suite, nous noterons

$$Sym(X) \stackrel{def}{=} \{f : X \rightarrow X, f \text{ inversible}\}$$

le groupe des applications inversibles (régulières ou non) d'un ensemble  $X$  dans lui même.

Au cours du temps, l'algorithme va être confronté à la loi sensorimotrice, et nous voudrions qu'il en extraie quelque chose d'intinsèque, i.e. qui ne dépende pas de la manière dont l'information est encodée. L'indépendance au codage sensorimoteur signifie que nous voudrions extraire les mêmes choses de la loi  $\varphi$  et de la loi  $f \circ \varphi \circ g$ , où  $(f, g) \in Sym(\mathcal{S}) \times Sym(\mathcal{M})$  est un simple changement de code des activités neuronales. La première question, avant de savoir comment il peut faire, est de se demander quelles structures invariantes peuvent exister. Considérons le groupe suivant :

$$\Gamma(\varphi) \stackrel{def}{=} \{g \in Sym(\mathcal{M} \times \mathcal{E}) \text{ tel que } \varphi \circ g = \varphi\}$$

**Proposition 1.** *La structure de groupe de  $\Gamma(\varphi)$  est invariante par changement de code sensorimoteur, et elle retient de  $\varphi$  tout ce qui est indépendant du code.*

Nous allons tout de suite donner un exemple concret : imaginons que les inputs sensoriels de notre algorithme soient constitués par, mettons, les images provenant de deux caméras observant le même ensemble d’objets, et que ses outputs moteurs lui permette de bouger ces deux caméras. Si ce contrôle est fonctionnel (au sens où la position instantanée des caméras est *fonction* des ordres moteurs instantanés, par opposition à non-holonomique), le résultat précédent nous indique que la *seule* chose qui ne dépende pas du codage dans cette loi sensorimotrice, c’est le groupe des transformations des ordres moteurs et de la configuration des objets laissant invariante la loi sensorimotrice.

Dans le cas simplifié où les outputs moteurs sont en relation univoque avec la position des caméras, on peut identifier ces transformations à des symétries de  $\mathcal{P} \times \mathcal{E}$  vis-à-vis du système sensoriel, et nous sommes naturellement intéressés par le sous-groupe connexe à l’identité de ces symétries puisqu’il est étroitement relié aux modifications continues des couples  $(P, E)$  : les mouvements. Or si les champs visuels des caméras se recoupent, du fait de la stéréoscopie les seuls mouvements laissant invariants les inputs sensoriels sont *les mouvements rigides* du système {caméras,objets} et l’on accède ainsi à une propriété “objective” de l’espace dans lequel se trouve les caméras. On note par ailleurs que le groupe des symétries associé agit indépendamment sur les caméras et sur les objets : pour un mouvement rigide  $g$  agissant sur le système {caméras,objets}, la configuration de l’environnement n’influence pas la position finale des caméras, et réciproquement.

Finalement, de cet exemple nous ne retiendrons que trois conditions abstraites : différentiabilité, connexité et action diagonale sur  $\mathcal{M} \times \mathcal{E}$ , qui vont s’avérer suffisante pour élaborer un algorithme capable d’extraire tout seul et avec une grande généralité, i.e. que ce soit le groupe euclidien ou non, la structure du sous-groupe de  $\Gamma(\varphi)$  des transformations de l’environnement associées aux mouvements du système {corps,environnement} qui laissent les inputs sensoriels invariants.

Notons  $\Gamma^d(\varphi)$  le sous-groupe des éléments de  $\Gamma(\varphi)$  agissant diagonalement sur  $\mathcal{M} \times \mathcal{E}$ , i.e. tels que  $g(M, E) = (g_1(M), g_2(E))$ , et  $\Gamma_0^d(\varphi)$  la composante connexe à l’identité de  $\Gamma^d(\varphi)$  (c’est un sous-groupe de  $\Gamma^d(\varphi)$ ). Par commodité, nous noterons  $\phi_1$  et  $\phi_2$  son action sur  $\mathcal{M}$  et sur  $\mathcal{E}$  ( $\phi_1(g, M) = g_1(M)$  et  $\phi_2(g, E) = g_2(E)$ ), ce qui donne l’équation :

$$\varphi(\phi_1(g, M), \phi_2(g, E)) = \varphi(M, E) \quad \forall (g, M, E) \in \Gamma_0^d(\varphi) \times \mathcal{M} \times \mathcal{E} \quad (1)$$

Alors, si le système sensoriel est assez riche vis-à-vis de l’environnement et les capacités de modification de la position assez grandes, la section 1.2 montre qu’il est possible de définir une action de  $\Gamma_0^d(\varphi)$  sur les inputs sensoriels dont les champs de vecteurs fondamentaux sont expérimentables par l’algorithme à partir d’une famille finie de vecteurs particuliers, la section 1.3 montre comment l’algorithme peut trouver tout seul cette famille, et la section 1.4 rappelle comment les champs de vecteurs fondamentaux révèlent une partie de la structure du groupe qui les a générés.

## 1.2 Champs de vecteurs fondamentaux sur la variété sensorielle

Nous allons supposer que les inputs sensoriels sont suffisamment riches pour que, lorsque le corps est immobile, aucun mouvement de l’environnement ne puisse être réalisé sans produire une modification quelconque des inputs sensoriels. En langage mathématique, nous supposons donc que :

**Condition 1.** *Il existe  $\{M_0\} \times \mathcal{V} \subset \mathcal{M} \times \mathcal{E}$  tel que  $\varphi(M_0, \cdot)$  soit un plongement de  $\mathcal{V}$  dans  $\mathcal{S}$*

Dans ce cas,  $\varphi(M_0, \mathcal{V})$  est une variété et  $\varphi(M_0, \cdot)$  un difféomorphisme de  $\mathcal{V}$  sur  $\varphi(M_0, \mathcal{V})$ . Nous écrivons  $\varphi^{-1}(M_0, \cdot)$  son inverse.

Pour le  $M_0 \in \mathcal{U}$  donné, il est alors possible de définir une action  $\phi^{M_0}$  de  $\Gamma_0^d(\varphi)$  sur  $\varphi(M_0, \mathcal{V})$  :

$$\phi^{M_0}(g, S) = \varphi(M_0, \phi_2(g, \varphi^{-1}(M_0, S))) \quad \forall S \in \varphi(M_0, \mathcal{V})$$

et d'associer à tout champ de vecteurs  $X$  invariant à gauche sur  $\Gamma_0^d(\varphi)$  un champ de vecteurs fondamental  $X^S$  sur  $\varphi(M_0, \mathcal{V})$  défini par :

$$X^S(S) = \frac{d}{dt} \phi^{M_0}(e^{-tX}, S)|_{t=0} \quad \forall S \in \varphi(M_0, \mathcal{V})$$

L'idée, sur laquelle on donne plus de détails dans la section 1.4, est que l'ensemble de ces champs munis du crochet de Lie sur les champs de vecteur, est isomorphe à l'algèbre de Lie d'un sous-groupe de  $\Gamma_0^d(\varphi)$  (celui qui a vraiment un effet sur les  $S$  au travers de l'action  $\phi^{M_0}$ ), et va donc nous permettre de comprendre ce sous-groupe d'invariants. Et le point crucial pour nous est que *la totalité du champ de vecteurs  $X^S$  (sur  $\varphi(M_0, \mathcal{V})$ ) peut être découvert expérimentalement à partir d'un seul vecteur*. Supposons en effet que l'algorithme connaisse le vecteur  $\frac{d}{dt} \phi_1(e^{-tX}, M_0)|_{t=0} \in T\mathcal{M}|_{M_0}$  (l'espace tangent à  $\mathcal{M}$  en  $M_0$ ), que nous appellerons un générateur du champ. Alors, il lui est possible de construire un ordre moteur  $M_X(t)$  vérifiant

$$\begin{cases} M_X(0) &= M_0 \\ \dot{M}_X(0) &= -\frac{d}{dt} \phi_1(e^{-tX}, M_0)|_{t=0} \end{cases}$$

et d'observer le champ de vecteurs fondamental en utilisant la propriété suivante :

**Proposition 2.**  $X^S(\varphi(M_0, E)) = \frac{d}{dt} \varphi(M_X(t), E)|_{t=0} \quad \forall E \in \mathcal{V}$

Lorsque l'environnement a bougé de manière à produire un certain input sensoriel, la valeur du champ de vecteurs fondamental en ce point est obtenu par l'application du mouvement générateur.

### 1.3 Découvertes des générateurs dans $T\mathcal{M}|_{M_0}$

**Proposition 3.** *S'il existe de plus  $E_0 \in \mathcal{V}$  tel que  $\varphi$  vérifie la propriété :*

$$\frac{\partial \varphi}{\partial M}(M_0, E_0) \dot{M} = 0 \Rightarrow \frac{\partial \varphi}{\partial M}(M_0, E) \dot{M} = 0 \quad \forall E \in \mathcal{V}$$

alors, pour tout ordre moteur  $M(t)$

$$\frac{d}{dt} \varphi(M(t), E_0) = \frac{d}{dt} \varphi(M_X(t), E_0) \Rightarrow \frac{d}{dt} \varphi(M(t), E) = X^S(\varphi(M_0, E)) \quad \forall E$$

Lorsque cette condition est vérifiée, le choix particulier d'un ordre moteur parmi ceux qui ont la même conséquence sensorielle tangente lorsque l'environnement est dans la configuration  $E_0$  n'a donc pas d'importance dans la construction du champ de vecteur fondamental. C'est le cas, comme nous allons le voir, lorsque par exemple il n'y a pas de changement de position  $P$  qui laisse invariant les inputs sensoriels (ce qui n'empêche pas qu'il puisse y avoir plusieurs outputs moteurs qui produisent la même position) :

**Proposition 4.** *Si  $\frac{\partial \varphi_b}{\partial P}(\varphi_a(M_0), E_0)$  est injective, alors la condition de la propriété précédente est vérifiée.*

La recherche des générateurs devient dans ce cas la recherche de leurs conséquences sensorielles tangentes. Or cette recherche là est limitée par la proposition 2, qui implique que ces conséquences se situent dans un sous-espace linéaire de l'intersection des espaces tangents à  $\varphi(M_0, \mathcal{V})$  et  $\varphi(\mathcal{U}, E_0)$ , où  $\mathcal{U}$  est un voisinage de  $E_0$  dans  $\mathcal{E}$  :

$$\forall X, X^S \in T\varphi(M_0, \mathcal{V})|_{S_0} \cap T\varphi(\mathcal{U}, E_0)|_{S_0}$$

Par ailleurs, la proposition suivante montre que c'est précisément les conséquences sensorielles vérifiant cette condition qu'il faut chercher :

**Proposition 5.** *Tout générateur moteur dont les conséquences sensorielles tangentes auraient, quelque soit la configuration de l'environnement, pu être produites par un mouvement de ce dernier, i.e. tel que*

$$\frac{d}{dt} \varphi(M(t), E) \in T\varphi(M_0, \mathcal{V})|_{\varphi(M_0, E)} \quad \forall E \in \mathcal{V}$$

*révèle un groupe connexe monodimensionnel indépendant du codage.*

#### 1.4 Algèbre de Lie de $\Gamma_0^d(\varphi)$

Les champs de vecteur fondamentaux sont la clef de notre problème en raison du morphisme (voir par exemple [6]) :

$$[X^S, Y^S] = [X, Y]^S$$

où le terme de gauche utilise le crochet de Lie des champs de vecteurs sur  $\varphi(M_0, \mathcal{V})$  (que l'on peut expérimenter), et le terme de droite utilise le crochet de Lie dans l'algèbre de Lie de  $\Gamma_0^d(\varphi)$  (que l'on cherche à comprendre). Le problème restant est de s'intéresser à l'efficacité de l'action  $\phi^{M_0}$ . Or

$$\phi^{M_0}(g_1, \cdot) = \phi^{M_0}(g_2, \cdot) \Leftrightarrow \phi_2(g_1, \cdot) = \phi_2(g_2, \cdot)$$

ce qui veut dire que pour une transformation  $g \in \Gamma_0^d(\varphi)$ , seule la partie de  $g$  qui agit sur l'environnement a un effet. On peut donc se restreindre à un sous-groupe pour lequel l'action est efficace, en considérant  $\Gamma_0^d(\varphi)/H$  où

$$H \stackrel{def}{=} \{g \in \Gamma_0^d(\varphi) \text{ tels que } \phi_2(g, \cdot) = Id_{\mathcal{E}}\}$$

est un sous-groupe distingué.

Dans ces conditions, on a alors les propriétés suivantes :

1.  $X \mapsto X^S$  est un morphisme d'algèbres de Lie injectif : la structure abstraite de l'algèbre de Lie du groupe est accessible au travers de son action.
2. tous les éléments du groupe (parce qu'il est connexe) peuvent s'écrire sous la forme  $g = e^{X_1} e^{X_2} \dots e^{X_k}$ , ces exponentielles sur le groupe étant mises en relation avec les actions sur les inputs sensoriels par

$$\phi^{M_0}(g, S) = \phi^{M_0}(e^{X_1}, \phi^{M_0}(e^{X_2}, \dots \phi^{M_0}(e^{X_k}, S)))$$

où l'action de l'exponentielle de  $tX$  est obtenue par le flot à l'instant  $t$  du champ  $X^S$  (qu'on sait construire).

On a donc trouvé comment accéder, à un isomorphisme près (ce qui était de toute façon ce qu'on peut faire de mieux puisque seule la structure est invariante par changement de code moteur), à l'algèbre de Lie de  $\Gamma_0^d(\varphi)/H$  et même à sa structure de groupe.

Il reste à le décomposer en fonction de ses propriétés algébriques, ce qui peut se faire dans un cadre général. Dans notre cas (celui du groupe euclidien) on voudrait identifier les éléments particuliers que sont les translations et les rotations ; ce qui est réalisable par un algorithme utilisant exclusivement des principes d'algèbres linéaire. Nous ne le présentons pas ici car son intérêt n'est que pratique, l'essentiel est le résultat : le crochet de Lie permet effectivement cette décomposition.

## 2 Simulations

Nous avons utilisé ces résultats pour l'étude des lois sensorimotrices générées par des corps virtuels satisfaisant les différentes conditions requises pour notre approche. La condition la plus critique, qui justifie l'emploi de simulations plutôt que de véritables robots, est l'aspect fonctionnel de la loi sensorimotrice qui permet les développements précédents. Pour des raisons de simplicité algorithmique, il semblait en effet indiqué de commencer par tester le cas où il existe une relation fonctionnelle instantanée entre les inputs sensoriels et outputs moteurs, avant de s'attaquer à une situation où il faut raisonner plus généralement en termes de trajectoires motrices et trajectoires sensorielles. Il s'agit de plus du cas sans doute le plus intuitif et donc le plus illustratif de nos idées.

### 2.1 Description du rat virtuel

Dans l'expérience du rat virtuel, la loi sensorimotrice connectée à notre algorithme est inspirée de celle qui pourrait-être impliquée par la tête d'un rat.

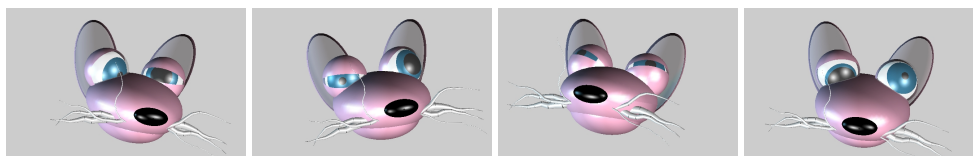


FIG. 1 – L'organisme virtuel utilisé pour les simulations. Des commandes motrices aléatoires produisent des modifications complexes et arbitraires du corps, impliquant des mouvements de la tête, des changements dans les directions de fixation des yeux, des changements d'ouverture des paupières et du diaphragme.

La tête virtuelle que nous avons simulée dispose de capteurs visuels, sensoriels et tactiles (voir la figure 2). Le dispositif visuel est constitué de deux yeux, chacun lui-même constitué de 40 cellules photosensibles distribuées aléatoirement sur une rétine bidimensionnelle, une lentille, un diaphragme et deux paupières. Les images des sources lumineuses constituant l'environnement sont projetées au travers de la lentille sur la rétine pour venir stimuler les cellules photoélectriques, avec un influx lumineux total déterminé par l'ouverture du diaphragme et des paupières. Le dispositif auditif est constitué d'un capteur d'intensité sonore dans chacune des deux oreilles, avec un profil de sensibilité qui favorise les sources auditives situées avec un azimut et une élévation nulle par rapport à l'orientation de la tête. Le dispositif tactile est constitué de 4 vibrisses sur chaque côté de la mâchoire, qui accrochent les objets touchés (jusqu'à une certaine limite) et délivrent un signal fonction de l'écart d'orientation à la position de repos. La stimulation de l'ensemble des capteurs est donc donné par un vecteur à  $2 \times (40 + 1 + 4) = 90$  coordonnées, et les inputs sensoriels sont ensuite produits par le mélange de toutes ses informations via une matrice arbitraire  $W_S \in \mathcal{M}(90, 300)$  représentant le code neuronal sensoriel. L'ensemble des inputs sensoriels est donc dans ce cas une sous-variété de  $\mathcal{R}^{300}$ .

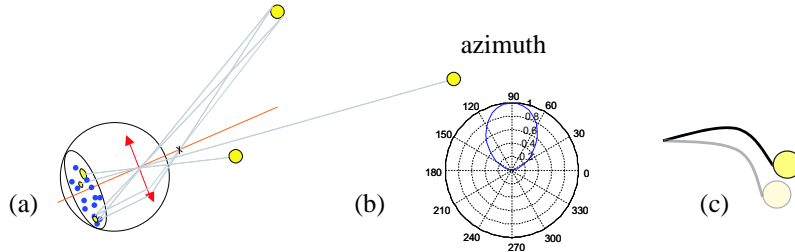


FIG. 2 – Le système sensoriel. (a) La partie sensorielle de chaque œil est constituée de cellules photosensibles distribuées de manière aléatoire sur la rétine (points foncés). (b) Les capteurs auditifs ont un profil de gain favorisant les sons émis par les sources situées devant la tête (c) Les capteurs tactiles sont pliés lorsqu'ils sont en contact avec une source.

Le système moteur était constitué comme suit. Seize paramètres de contrôle étaient extraits à partir de combinaisons linéaires des outputs moteurs en utilisant une matrice arbitraire  $W_M \in \mathcal{M}(16, 300)$  représentant le code neuronal moteur. La configuration de la tête du rat était ensuite calculée en utilisant les six premiers paramètres pour fixer la position et l'orientation de la tête, les cinq suivants pour déterminer l'orientation et l'ouverture du diaphragme et des paupières de l'œil gauche, et les cinq derniers pour déterminer l'orientation et l'ouverture du diaphragme et des paupières de l'œil droit. Les vibrisses n'étaient pas contrôlables indépendamment de la tête.

Dans la simulation nous avons utilisé des encodages linéaires  $W_S$  et  $W_M$  afin de montrer que l'algorithme n'avait pas besoin de connaître le code et que la dimension des inputs sensoriels et des outputs moteurs pouvait être très élevée. Mais il faut bien noter que n'importe quelle fonction non-linéaire aurait pu tout aussi bien être utilisée ; la loi sensorimotrice n'est de toute façon absolument pas linéaire, même dans le cas où les encodages le sont : les capteurs délivrent en effet un signal qui n'est pas linéaire par rapport à la configuration de la tête, et celle-ci n'est pas linéaire par rapport aux ordres moteurs.

## 2.2 Estimation de l'espace tangent

L'objectif de notre algorithme est d'estimer les conséquences sensorielles tangentes des ordres moteurs, et d'en déduire les générateurs moteurs possibles. Comme on l'a démontré, cela passe par l'estimation de l'intersection des espaces tangents à la variété  $\varphi(M_0, \mathcal{V})$  observée lorsque l'on donne un output moteur constant, et à la variété  $\varphi(\mathcal{U}, E_0)$  observée lorsque l'environnement est fixe et que l'on donne différents ordres moteurs. Nous nous limiterons ici à cette première étape que constitue l'extraction non-supervisée des générateurs, mais l'intérêt de ce qui précède est bien sûr de montrer comment ces vecteurs, attachés à un seul point de l'espace moteur, sont associés à des structures étendues de manière précise sur tout un voisinage.

La méthode utilisée dans [8] pour estimer l'espace tangent, et plus spécifiquement sa dimension, nécessitait un niveau de précision problématique ( $10^{-8}$ ). L'une des raisons était l'instabilité de la décomposition en valeurs singulières (SVD) lorsque l'on a affaire à des matrices singulières (ce qui est bien sûr le cas, puisqu'on part du principe qu'il y a de la redondance et donc des applications non injectives). Nous avons développé une méthode beaucoup plus robuste et stable, qui utilise de plus une manière plus réaliste d'estimer la différentielle, via des dérivées temporelles et non plus une approximation linéaire locale.

L'idée est de bootstrapper la décomposition en valeurs singulières, c'est-à-dire que l'on va chercher à trouver les "bons" mouvements et les "bonnes" informations sensorielles tels que la relation entre leurs

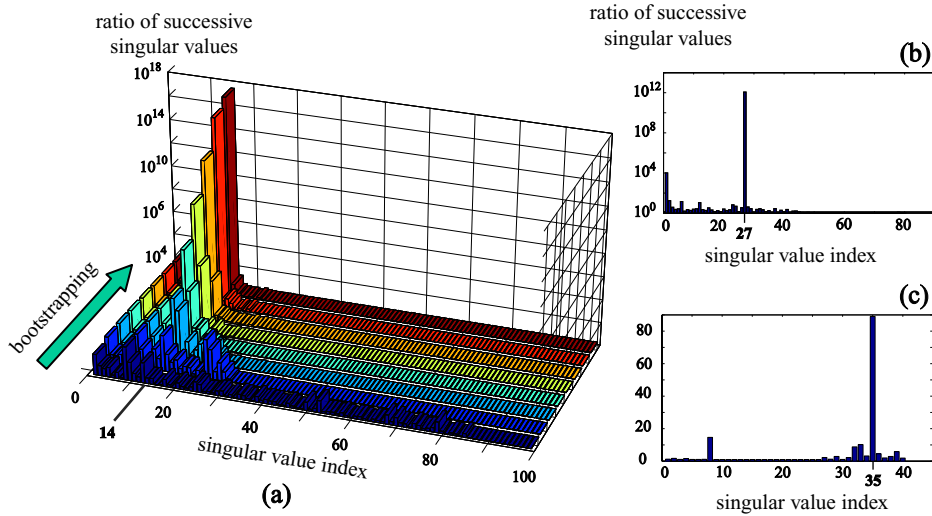


FIG. 3 – Estimation des conséquences sensorielles tangentes des mouvements générateurs. (a) l’estimation du rapport des valeurs singulières de la loi sensorimotrice tangente (quand  $E$  est fixe à  $E_0$ ) durant le processus de bootstrapping ; (b) la matrice correspondant à l’estimation d’une famille génératrice de l’espace tangent aux inputs sensoriels observés lorsque  $M$  est fixe à  $M_0$  ; (c) la matrice constituée par la réunion des deux familles de vecteurs trouvées dans les deux cas précédents. Les noyaux des deux premières matrices reflètent la redondance des variables ; le noyau de la dernière est reliée à l’intersection des deux plans tangents. Le graphique montre qu’il y a 14 variables de contrôle du corps et 27 pour l’environnement (voir le texte). Le noyau de la dernière matrice conduit au calcul d’une intersection de dimension 6, reflétant l’espace tangent au groupe euclidien  $SE(3)$  (voir le texte).

dérivées soit diagonale : la SVD est utilisée à chaque étape pour extraire et préciser des bonnes directions, mais elle n’est jamais utilisée comme résultat final ; ce dernier est obtenu comme la base des bons mouvements et des bonnes informations lorsqu’on estime que la différentielle dans cette base est assez diagonale. Ainsi, les erreurs dues à l’estimation des dérivées temporelles (deux chiffres significatifs seulement) ne sont pas amplifiées par la SVD. Il faut noter que les erreurs d’estimation des dérivées temporelles sont de plus dûes en majeure partie à la simulation numérique, la dérivée d’une fonction discrète étant un problème mal formulé. Dans l’optique d’une implémentation réaliste, un dérivateur physique travaillant en analogique serait beaucoup plus efficace : il faut échantillonner la dérivée, et pas dériver l’échantillonnage numérique.

Quoi qu’il en soit, la figure 3 présente l’évolution de l’estimation des dimensions de l’espace tangent au cours du bootstrapping. Il est clair que se limiter à la première étape ne permet pas d’identifier de manière fiable une différence d’ordre de grandeur entre les valeurs singulières, mais que celle-ci émerge de manière indiscutable au cours des estimations réalisées dans des directions de mieux en mieux adaptées.

En utilisant cette méthode comme première étape de l’expérience, lorsque l’environnement est fixe, l’algorithme calibre les signaux provenant de sa tête en même temps qu’il trouve une base pour l’espace tangent : il extrait les combinaisons linéaires de ses inputs sensoriels qui sont significatives au regard de sa propre mobilité. Durant la seconde étape, l’algorithme estime l’espace tangents aux inputs sensoriels observés lorsque l’environnement bouge (en utilisant les combinaisons précédemment obtenues comme base des inputs sensoriels “intéressants”). Finalement, il calcule l’intersection des deux espaces vectoriels précédents.

Pour terminer, l'algorithme choisit au moyen de la pseudo-inverse (qui est facile à calculer puisqu'on est dans une base où la loi sensorimotrice tangente est diagonale) une base de l'espace moteur tangent dont les conséquences sensorielles sont dans cette intersection et constituent donc un nombre fini de candidats au statut de générateur moteur.



FIG. 4 – Les effets des commandes motrices correspondant à une base génératrice extraite par l'algorithme. Elles révèlent le contrôle de la tête d'une manière rigide (à comparer à la figure 1). Sans le crochet de Lie pour traduire la commutativité, ces mouvements impliquent une composition arbitraire entre translations et rotations, et l'incommensurabilité de ces différentes transformations.

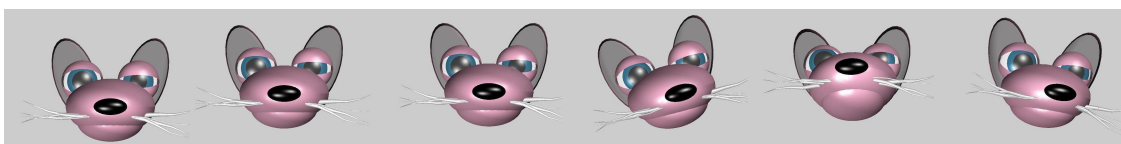


FIG. 5 – Les effets des commandes motrices extraites après décomposition du crochet de Lie : base orthogonale de translations (premières images), base d'un sous-groupe de rotations pures (dernières images). Outre la distinction translations/rotations, le crochet de Lie permet de normaliser l'amplitude de toutes les commandes vis à vis des transformations géométriques de la tête.

## 2.3 Résultats

La figure 3a montre l'évolution de l'estimation du rapport entre deux valeurs singulières successives. Le maximum de ce rapport révèle en effet la frontière entre deux groupes de valeurs plus comparables entre elles qu'entre les deux groupes, ce qui permet d'estimer le nombre de valeurs significativement non nulles. L'algorithme découvre ainsi qu'il faut 14 paramètres pour caractériser la configuration du corps (sur les 16 utilisés, 2 se combinent pour pour chaque oeil puisque l'ouverture du diaphragme ou des paupières ont vis-à-vis des inputs sensoriels exactement le même effet dans notre modèle, c'est-à-dire la modification de l'influx lumineux total). Il trouve ensuite qu'il faut 27 variables pour décrire l'environnement (neuf lampes ponctuelles dans un espace tridimensionnel), figure 3b. Le calcul de l'intersection entre les deux espaces tangents estimés fait apparaître un espace de dimension  $14 + 27 - 35 = 6$ . Les générateurs moteurs extraits par l'algorithme ont pour conséquence les mouvements présentés dans la figure 4, à comparer avec les mouvements aléatoires de la figure 1. Finalement, l'estimation du crochet de Lie (dont le détail n'est pas expliqué ici par manque de place) et sa décomposition (idem) permet de distinguer entre les générateurs des translations et les générateurs des rotations pures, ainsi que d'établir une commensurabilité entre eux. Les mouvements résultant sont présentés dans la figure 5.

### 3 Conclusion

La naïveté de notre modèle quant à son réalisme physique donne cependant l'espoir d'une formulation précise de la problématique de la perception dans un cadre sensorimoteur, qui permette de répondre à la question centrale soulevée en introduction : qu'existe-t-il dans la loi sensorimotrice qui permette de placer *la nature physique* du stimulus au coeur de la perception, plutôt que se baser sur l'identité des aires neuronales traitant l'information ? Qu'existe-t-il, précisément, dans une loi sensorimotrice au code arbitraire, qui permette la réorganisation du cortex auditif en cortex visuel [9], ou la perception visuelle à partir des capteurs d'une autre modalité sensorielle [1, 2] ?

Comme premier élément de réponse, nous avons montré dans le cadre exposé comment il était possible, au travers de l'expérimentation d'une loi sensorimotrice impliquant des codes sensoriels et moteurs inconnus, d'extraire un ensemble de propriétés de cette loi qui sont indépendantes de ces codes et reflètent donc des propriétés "objectives" ou externes, dans le sens où les organismes ayant des inputs sensoriels aussi informatifs les uns que les autres mais dans des codes différents vont aboutir à la même structure. Le résultat est donc une perception à la fois incarnée, puisqu'elle dépend de l'expérience sensorimotrice et s'exprime dans le code sensorimoteur spécifique de l'organisme, mais basée sur des propriétés de la loi sensorimotrice qui ne dépendent pas de ce code.

Ceci est à mettre en parallèle avec les expériences de substitution sensorielle, où les sujets sont confrontés à une relation entre des stimuli tactiles et des actions motrices, artificiellement construite à partir de la loi sensorimotrice reliant habituellement les stimuli visuels et les actions motrices des yeux et de la tête. Les codes sont inconnus, mais certaines *propriétés* de la liaison semblent permettre au cerveau de réinterpréter la stimulation tactile pour l'extérioriser (c'est-à-dire la projeter dans un cadre spatial dépassant les frontières du corps) et retrouver ainsi une perception plus proche du visuel que du toucher [1].

### Références

- [1] P. Bach-y Rita, C. C. Collins, F. Sauders, B. White, and L. Scadden. Vision substitution by tactile image projection. *Nature*, 221 :963–964, 1969.
- [2] P. Bach-y Rita, K. Kaczmarek, M. Tyler, and J. Garcia-Lara. Form perception with a 49-point electro-tactile stimulus array on the tongue. *Journal of Rehabilitation Research Development*, 35 :427–430, 1998.
- [3] A. Berthoz. *Le sens du mouvement*. Odile Jacob, Paris, 1997.
- [4] H.V.B. Hirsch and D.N. Spinelli. Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, 168 :869–871, 1970.
- [5] M. Jeannerod. *Le cerveau Intime*. Odile Jacob, 2002.
- [6] T. Masson. *Géométrie différentielle, groupes et algèbres de Lie, fibrés et connexions*. LPT, 2001.
- [7] J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 2001.
- [8] D. Philipona, K. O'Regan, and J.-P. Nadal. Is there something out there ? Inferring space from sensorimotor dependencies. *Neural Computation*, 15(9), 2003.
- [9] J. Sharma, A. Angelucci, and M. Sur. Induction of visual orientation modules in auditory cortex. *Nature*, 404 :841–847, 2000.
- [10] L. von Melschner, S. L. Pallas, and M. Sur. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, 404 :871–876, 2000.