



## Un (petit) pas vers la perception interactive

Sylvain ARGENTIERI  
Habilitation à diriger des recherches

6 décembre 2018

David FILLIAT

François MICHAUD

Mathias QUOY

Philippe BIDAUD

Philippe SOUÈRES

Bruno GAS

Professeur à l'ENSTA Paritech

Professeur à l'Université de Sherbrooke, Canada

Professeur à l'Université de Cergy-Pontoise

Professeur à Sorbonne Université

Directeur de Recherche au LAAS-CNRS

Professeur à Sorbonne Université

Rapporteur

Rapporteur

Rapporteur

Examineur

Examineur

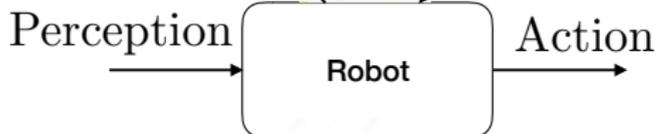
Référent HDR

# Introduction

Contexte :



Raisonnement



Autonomie

De l'automate ...

- Canard de Jacques de Vaucanson
- Véhicules de Braitenberg

... au robot :

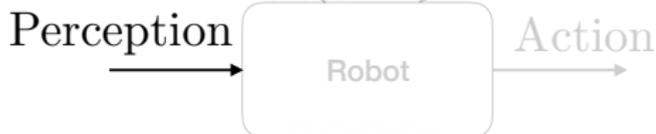
- doté de capacités de raisonnement, de décision
- doté de capacités d'adaptation et d'autonomie

# Introduction

Contexte :



Raisonnement



Autonomie

De l'automate ...

- Canard de Jacques de Vaucanson
- Véhicules de Braitenberg

... au robot :

- doté de capacités de raisonnement, de décision
- doté de capacités d'adaptation et d'autonomie

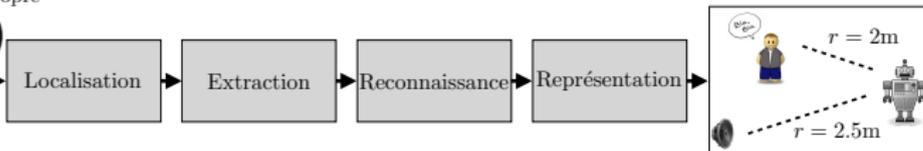
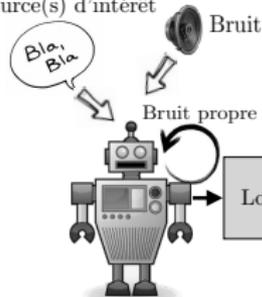
Perception :

- reproduisant les capacités sensorielles de l'Homme
- mais non limitée par elles

# Audition robotique

**Objectif** : analyser une scène sonore, dans un contexte robotique réaliste.

Source(s) d'intérêt



→ Problématiques largement traitées en Acoustique et Traitement du Signal ...

... abordée en Robotique selon deux paradigmes :

Antenne de microphones



AIST Project "Development of Humanoid Robot Type Intelligence Booster Platform", Japan, 2005

Audition binaurale



(KIM, NAKADAI et OKUNO, 2015)

# Audition robotique

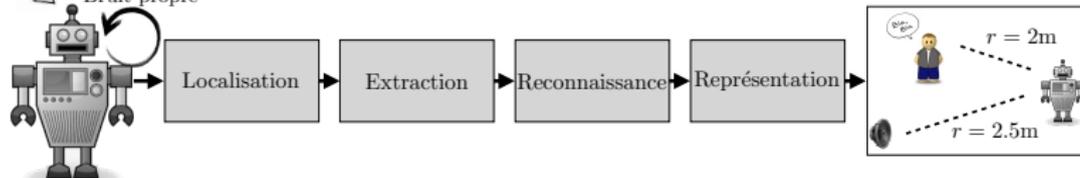
**Objectif** : analyser une scène sonore, dans un contexte robotique réaliste.

Source(s) d'intérêt

Bruit

Bla, Bla

Bruit propre



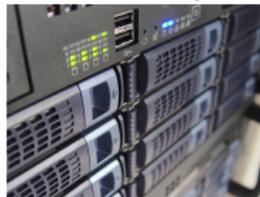
→ Problématiques largement traitées en Acoustique et Traitement du Signal ...

... mais le contexte robotique interroge les travaux de ces Communautés via de nouvelles :

Contraintes géométriques



Contraintes temporelles



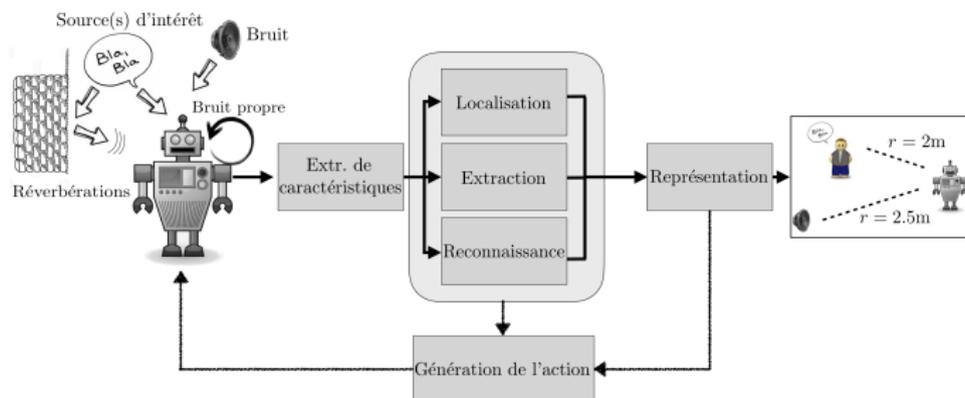
Contraintes environnementales



... tout en permettant de bénéficier d'une possibilité inédite : le **mouvement**.

# Audition robotique **active**

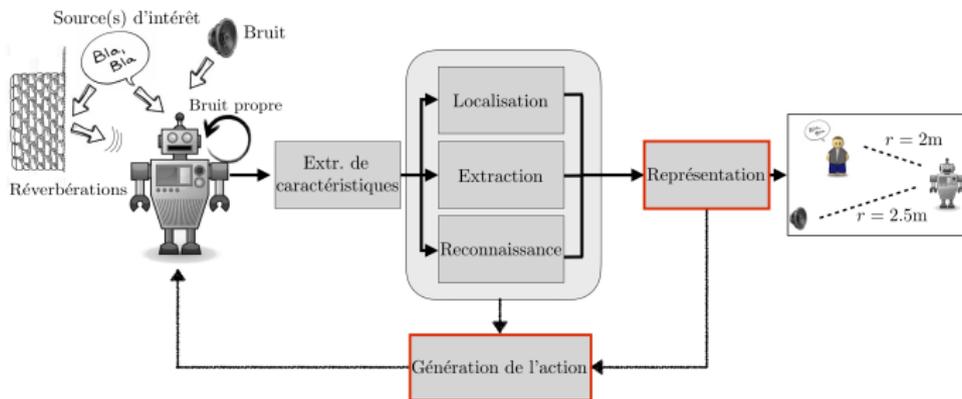
**Positionnement** : analyser une scène sonore, dans un contexte robotique **actif** réaliste.



Potential de l'approche active identifié depuis le début des années 2000 (NAKADAI et al., 2000), mais encore peu abordé (LU et COOKE, 2010 ; V. NGUYEN et al., 2017).

# Audition robotique **active**

**Positionnement** : analyser une scène sonore, dans un contexte robotique **actif** réaliste.



Problématique abordée progressivement à l'occasion de deux projets internationaux



BINaural Active Audition for Humanoid Robots  
(ANR/JST, 2010-2014)



TWO!EARS (FP7 FET Open, 2013-2017)

... et dans un contexte **binaural**

# Question

Comment l'action est-elle susceptible de contraindre, ou d'aider, l'analyse d'une scène sonore dans des environnements robotiques réalistes ?

Éléments de réponse :

- Prise en compte des réverbérations : étude systématique des indices binauraux
- Caractérisation des conséquences des changements de position dans l'environnement : apprentissage de la localisation

→ Thèse de Karim YOUSSEF (09/2010–10/2013)

- Approche intrinsèquement active de la localisation de source : estimation en ligne de l'azimut et distance d'une source sonore

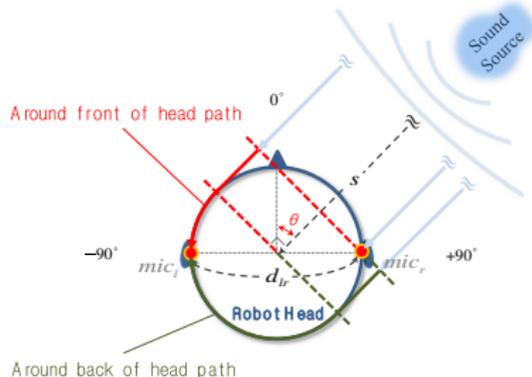
→ Thèse d'Alban PORTELLO (09/2010–12/2013)

- Vers des considérations attentionnelles : modulation du mouvement de la tête

→ Thèse de Benjamin COHEN-LHYVER (12/2013–09/2017)

# Comment localiser un son (en binaural) ?

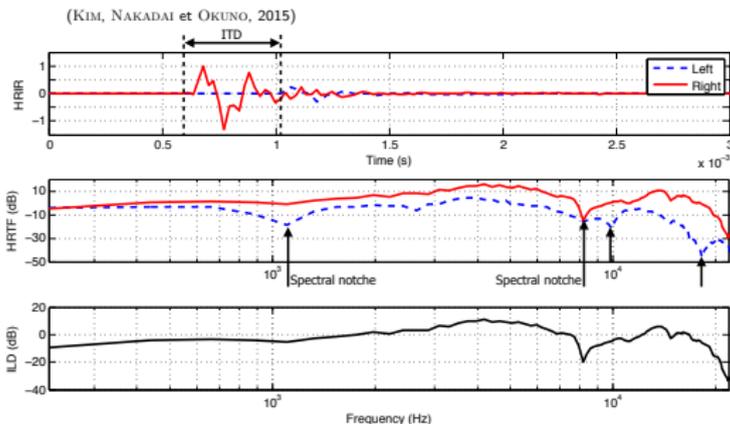
Propagation (dans le plan horizontal) :



- Différence interaurale en temps (ITD)
- Différence interaurale en phase (IPD)
- Différence interaurale en amplitude (ILD)

Head Related Transfer Function (HRTF) :

$$\begin{cases} L[k] = H_l(f[k], r_s, \theta_s, \varphi_s) S_s[k] \\ R[k] = H_r(f[k], r_s, \theta_s, \varphi_s) S_s[k] \end{cases}$$



(ARGENTIERI, DAVIES et SOUÈNES, 2015)

# Indices binauraux en robotique (1/2)

Un exemple (largement incomplet !) de ce qu'on trouve dans la littérature :

$$h(t) = at^{n-1}e^{-2\pi Bt} \cos(2\pi f_0 t + \phi)$$

$$ITD_{CC} = \frac{1}{f_e} \arg \max_m C_{lr}[m], \text{ avec } C_{lr}[m] = \sum_{n=0}^{N-m-1} l[n+m]r[n]$$

$$ITD_{AFE}^{(i)} = 10 \log_{10} \frac{\sum_{n=0}^{N-1} AFE(l^{(i)}[n])^2}{\sum_{n=0}^{N-1} AFE(r^{(i)}[n])^2}$$

$$ITD_{AFE}^{(i)} = \arg \max_m C_{lr}^{(i)}[m], \text{ avec } C_{lr}^{(i)}[m] = \sum_{n=0}^{N-m-1} AFE(l^{(i)}[n+m])AFE(r^{(i)}[n])$$

$$ILD = 10 \log_{10} \frac{\sum_{n=0}^{N-1} l[n]^2}{\sum_{n=0}^{N-1} r[n]^2}$$

$$G_{\text{PhaT}}[k] = \frac{1}{|L[k]||R[k]|}$$

$$ILD_{\text{FFT}}[k] = 20 \log_{10} \frac{|L[k]|}{|R[k]|}$$

$$ILD_{\text{FFT}_2}^{(i)} = 20 \log_{10} \frac{|\sum_{k=k_l[i]}^{k_u[i]} L[k]|}{|\sum_{k=k_l[i]}^{k_u[i]} R[k]|}$$

$$ITD_{\text{GCC}} = \frac{1}{f_e} \arg \max_m GCC_{lr}[m], \text{ avec } GCC_{lr}[m] = \text{IFFT}(G[k]L[k]R^*[k])$$

$$IPD_{\text{FFT}_1}^{(i)} = \sum_{k=k_l[i]}^{k_u[i]} IPD_{\text{FFT}}[k]$$

$$IPD_{\text{FFT}_1}^{(i)} = \sum_{k=k_l[i]}^{k_u[i]} ILD_{\text{FFT}}[k]$$

$$IPD_{\text{FFT}_2}^{(i)} = \arg \left( \sum_{k=k_l[i]}^{k_u[i]} L[k] \right) - \arg \left( \sum_{k=k_l[i]}^{k_u[i]} R[k] \right)$$

$$IPD_{\text{FFT}}[k] = \arg(L[k]) - \arg(R[k])$$

→ S'il y a consensus sur le sens (physique) des indices binauraux, il n'en est pas de même de leur estimation.

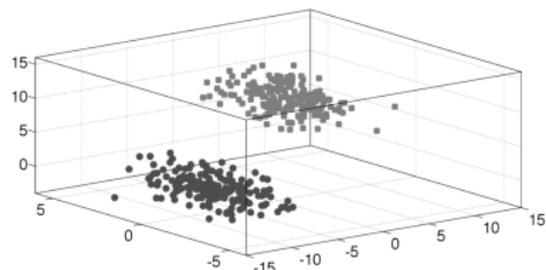
Comment sélectionner le "bon" indice, et selon quel critère ?

→ robustesse aux réverbérations

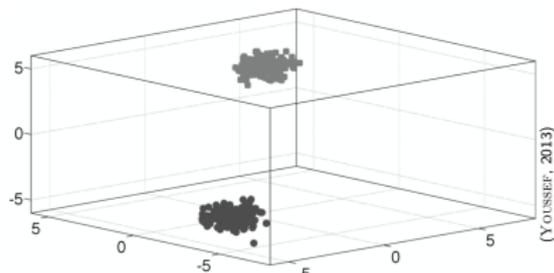
# Indices binauraux en robotique (2/2)

Constatation : tous les indices ne sont pas égaux devant les réverbérations

→ Critère de comparaison : Lambda de Wilks  $\Lambda$



$$\Lambda = 0.034$$



$$\Lambda = 0.006$$

(YOUSSEF, 2013)

Indices azimutaux :

- indices en amplitude plus “robustes”
- décomposition spectrale indispensable
- cochléaire > Fourier

Azimut

ITD et ILD après filtrage gammatone

Indices de distance :

- mauvais comportement en anéchoïque
- corrélation spatiale plus robuste

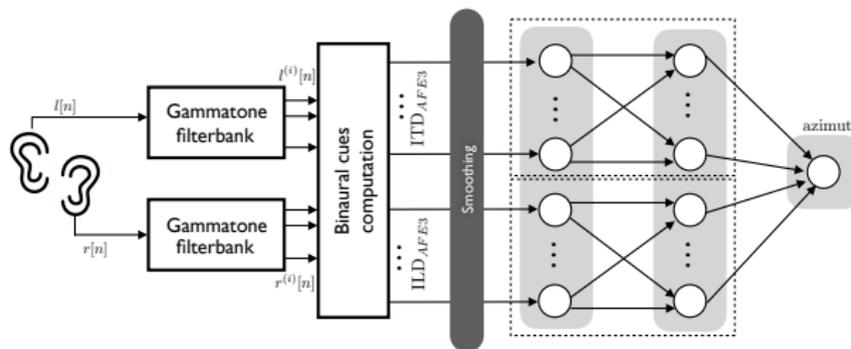
Distance

Calcul du DRR

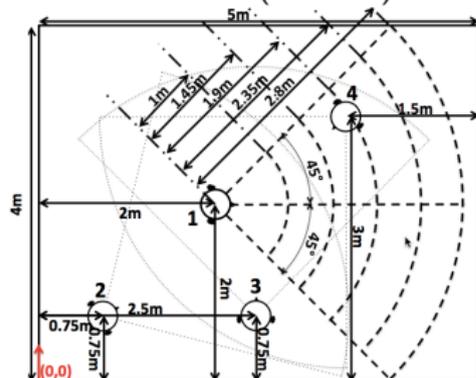
(K. YOUSSEF, S. ARGENTIERI et J. L. ZARADER, 2012)

# Apprentissage de la localisation : approche proposée

Pour l'azimut : réseau de neurones partiellement connecté



Base de données (simulées) :



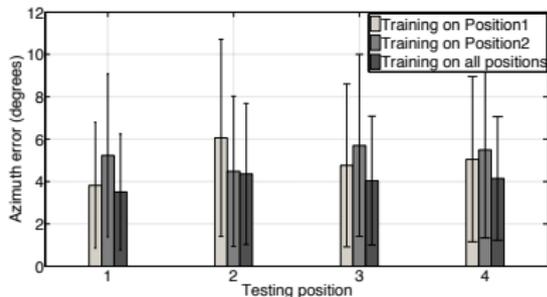
- 4 positions différentes du récepteur binaural (KEMAR) dans une pièce échoïque
- 19 valeurs d'azimut, 5 valeurs de distance, pour chaque position
- 4 valeurs différentes du  $RT_{60}$
- Source : parole ou bruit blanc

→ environ 9000 exemples

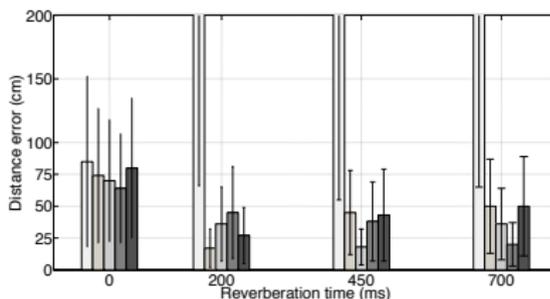
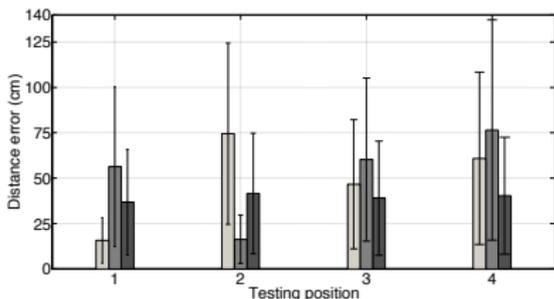
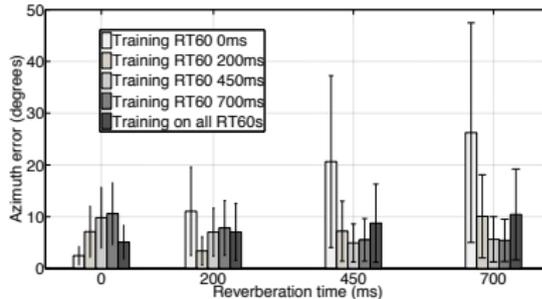
# Apprentissage de la localisation : résultats

... estimation de l'influence croisée des différents paramètres. Par exemple :

## Influence de la position ( $RT_{60} = 200\text{ms}$ )



## Influence du $RT_{60}$ (en position 1)



- position de l'apprentissage importante
- certaines positions sont difficiles

- conditions anéchoïques sans intérêt
- apprentissage multiconditionnel nécessaire

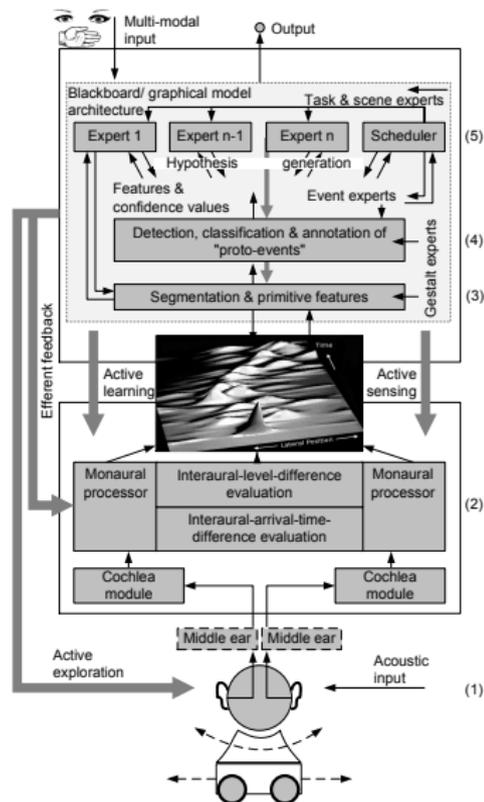
# Audition active : on ferme la boucle !



- **Objectif** : modèle computationnel de la perception sonore
- **Défi** : prise en compte de la multimodalité et des aspects actifs
- Approche mixte montante et descendante

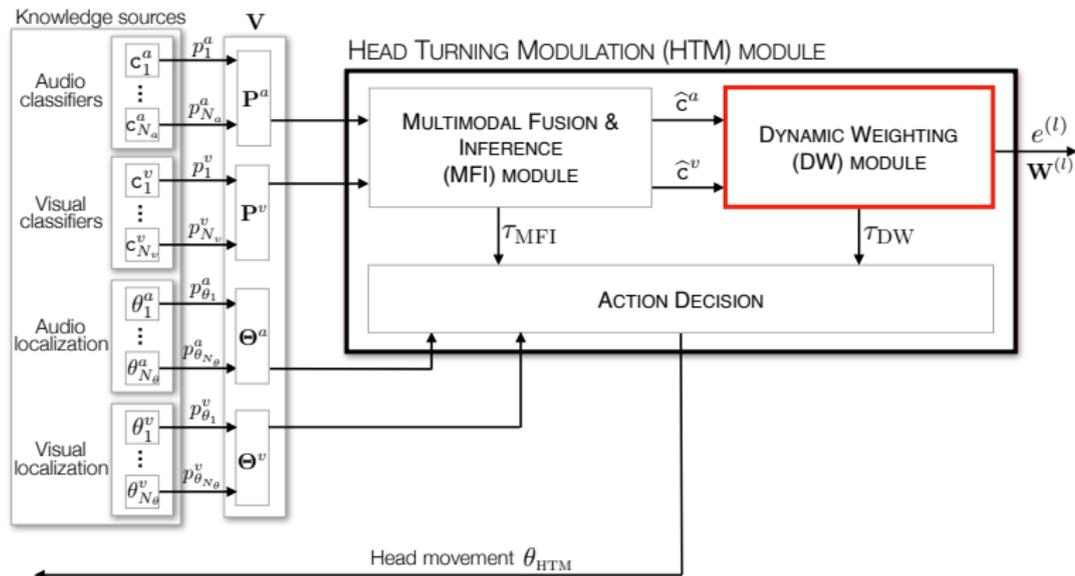
## Contribution :

- modulation du mouvement de la tête via des considérations attentionnelles
- exploitation du retour moteur pour la construction d'une représentation interne de l'environnement



# Le module “Head Turning Modulation” (HTM)

Architecture du système proposé :



Le module MFI :

- fournit les classes audio et visuelle  $\hat{c}^a$  et  $\hat{c}^v$
- même si les données sont manquantes et en présence d'erreurs

Le module DW :

- s'appuie sur la notion de congruence
- décide de l'importance d'un objet audiovisuel

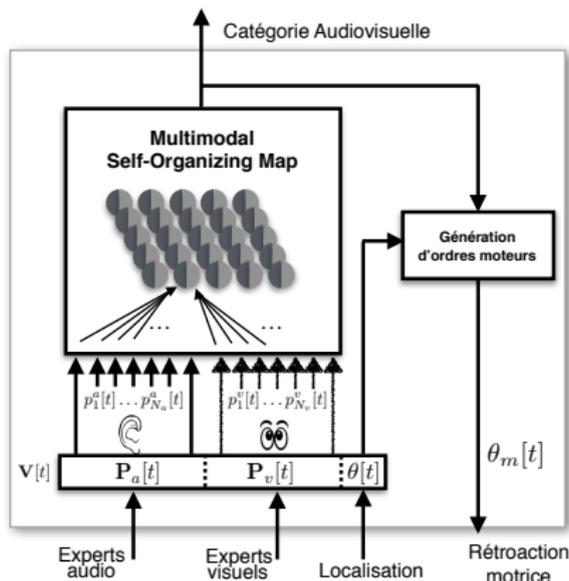
# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

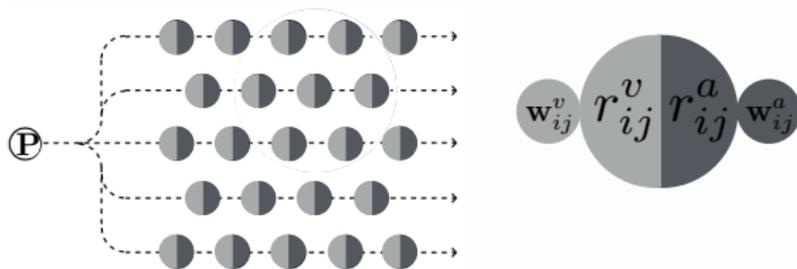
→ Architecture générale du module MFI :



# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

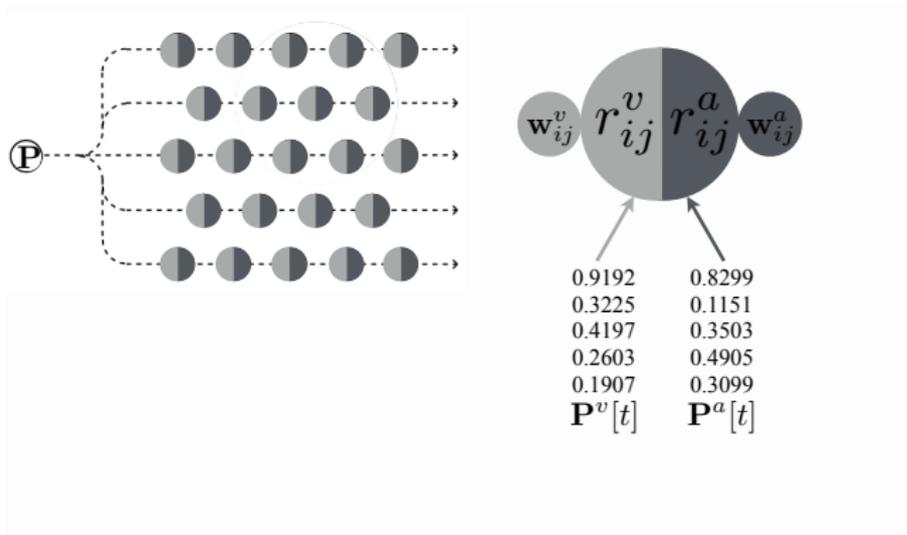
→ **M-SOM** : 2 poids  $w_{i,j}^v$  et  $w_{i,j}^a$  associé au nœud  $i,j$



# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

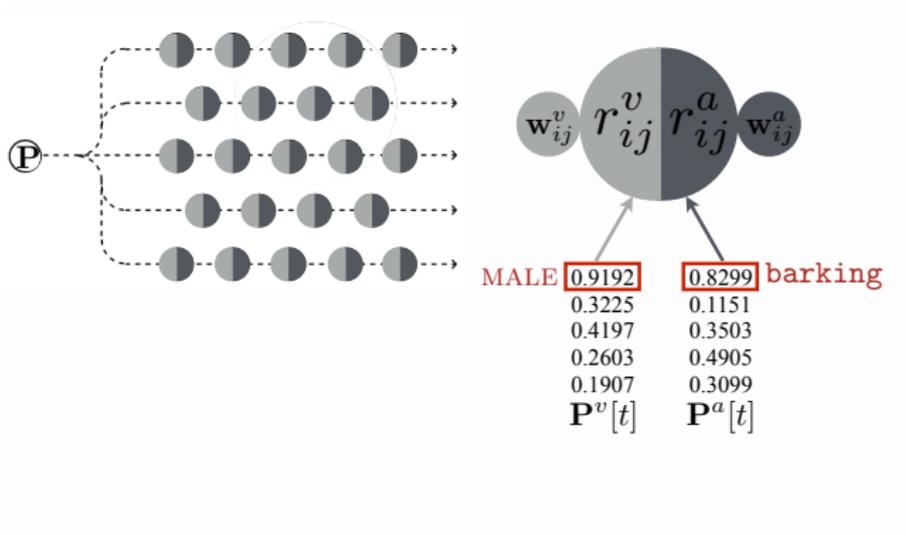
→ **M-SOM** : si toutes les données sont disponibles



# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si toutes les données sont disponibles

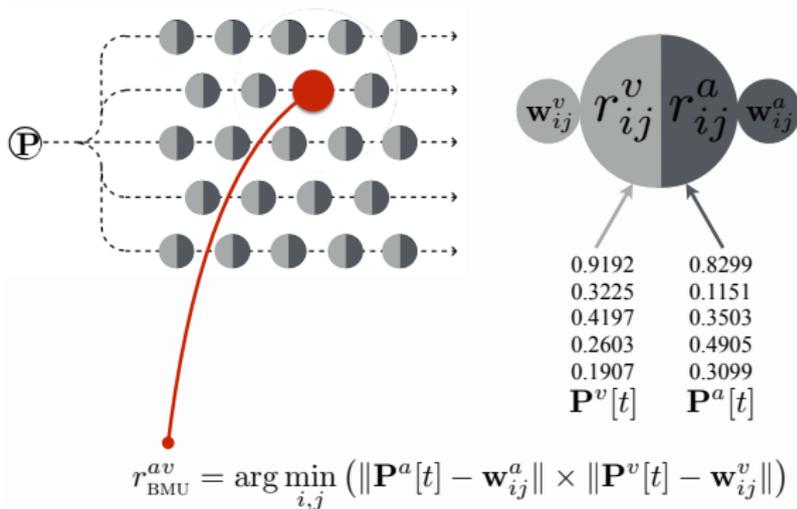


→ **Erreur de classification** en sortie de l'expert audio

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si toutes les données sont disponibles

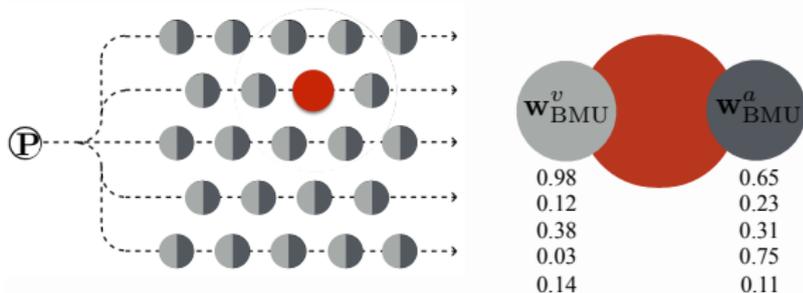


→ Détermination du nœud vainqueur

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si toutes les données sont disponibles

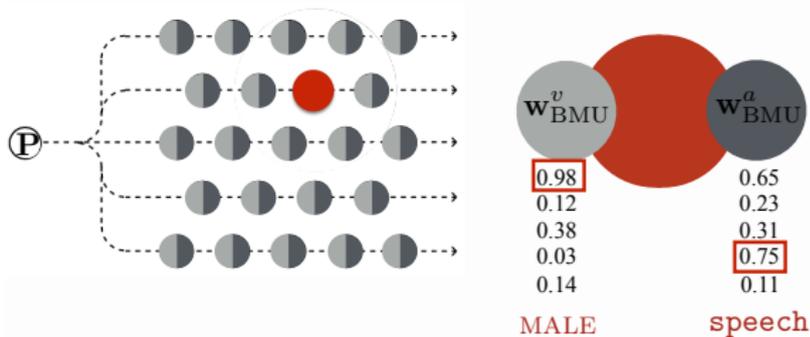


→ **Estimation de la classe audiovisuelle** et assimilation des données

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si toutes les données sont disponibles

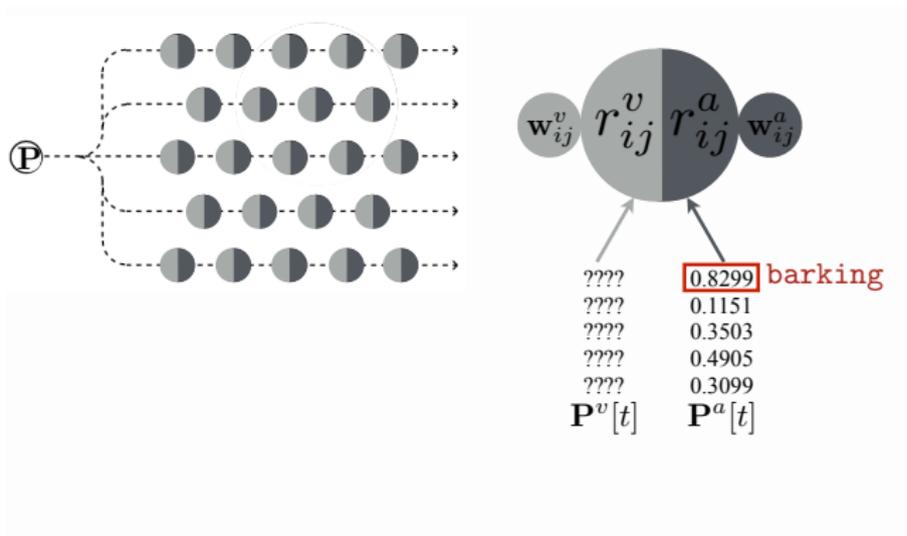


→ **Estimation de la classe audiovisuelle** et assimilation des données

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

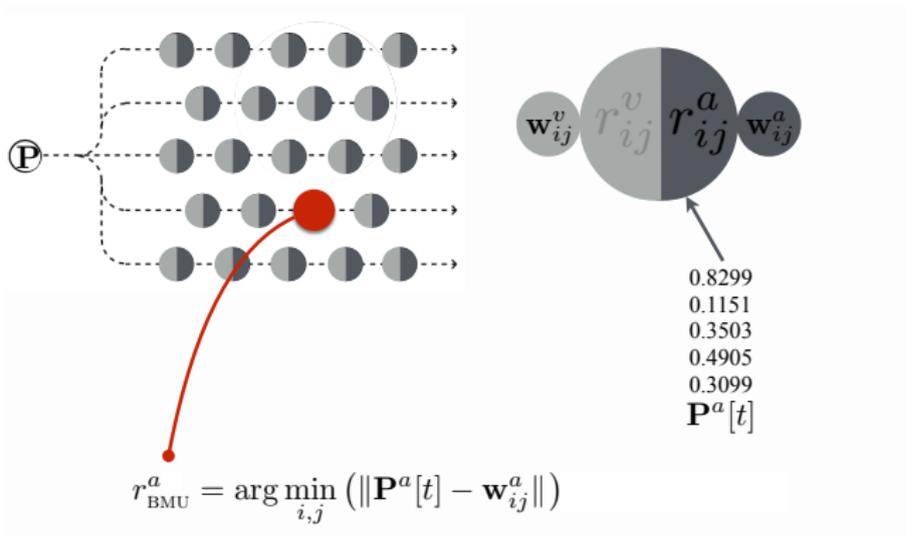
→ **M-SOM** : si les données visuelles sont manquantes



# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si les données visuelles sont manquantes

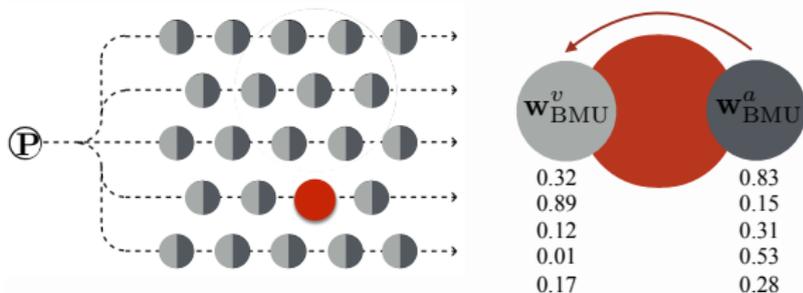


→ Détermination du nœud vainqueur, **sur la base de l'audio seulement**

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si les données visuelles sont manquantes

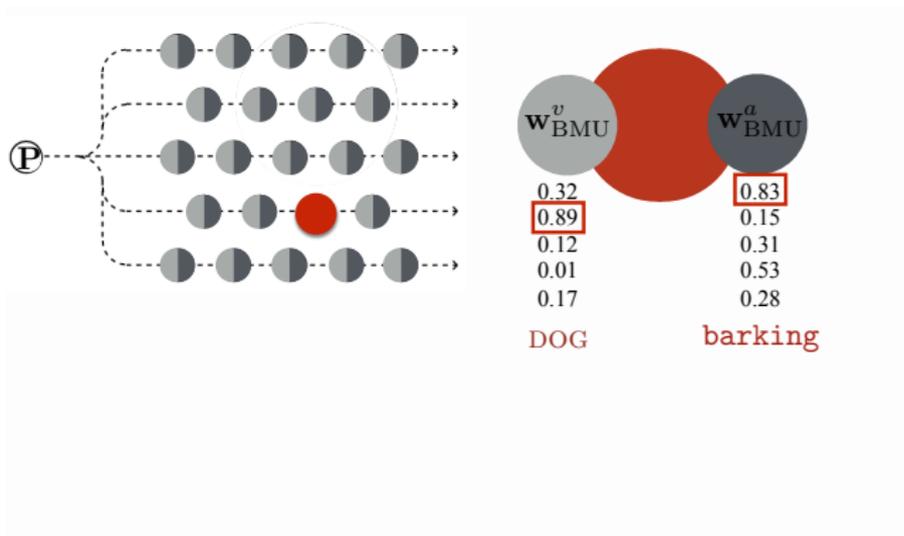


→ Comme chaque nœud dispose de deux vecteurs de poids, **on peut estimer la classe visuelle manquante**

# Inférence active de données audiovisuelles

**Question** : comment estimer la catégorie audio et visuelle d'une trame multimodale, même en présence d'erreurs de classification et d'absence de données ?

→ **M-SOM** : si les données visuelles sont manquantes



→ Pour s'assurer que l'inférence est correcte, le système peut provoquer un mouvement de tête en direction de l'évènement

# Modulation du mouvement de la tête et résultats

→ Définition d'un ratio d'inférence  $q$  permettant d'estimer la qualité de l'inférence, **et donc de moduler le mouvement de la tête** via un seuil  $K_q$  sur  $q$

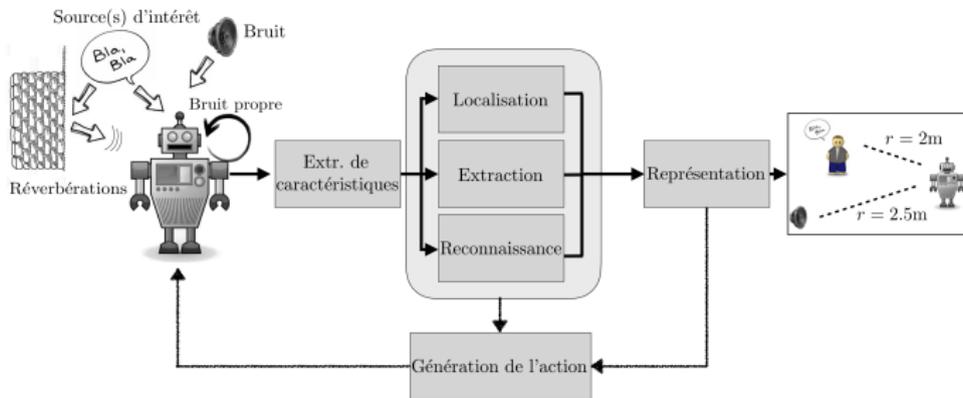
$$q\left(c^{(l)}(c_i^a, c_k^v)\right)[t] = \frac{\sum_1^t \delta_{ik}^{\text{miss}}[t-1] \delta_{ik}^{\text{all}}[t]}{\sum_1^t \delta_{ik}^{\text{miss}}[t]}, \text{ avec } \delta_{ik}^{\text{all/miss}}[t] = \begin{cases} 1 & \text{si } \hat{c}^{\text{all/miss}}(o_j) = \{c_i^a, c_k^v\}, \\ 0 & \text{sinon.} \end{cases}$$

→ Taux de bonne catégorisation audiovisuelle des évènements ( $\epsilon$  = taux d'erreurs des experts)

$\epsilon$	$N_I$	MFI	Robot omniscient	Robot naïf
0.3	3	0.992 (0.020)	0.703 (0.042)	0.414 (0.055)
	5	0.987 (0.022)	0.692 (0.017)	0.265 (0.014)
	7	0.942 (0.028)	0.691 (0.014)	0.198 (0.017)
	10	0.883 (0.041)	0.689 (0.011)	0.0145 (0.014)
	<b>moy</b>	<b>0.951</b>	<b>0.693</b>	<b>0.255</b>
0.7	3	0.774 (0.087)	0.282 (0.030)	0.165 (0.028)
	5	0.737 (0.105)	0.294 (0.014)	0.120 (0.023)
	7	0.683 (0.133)	0.296 (0.016)	0.081 (0.012)
	10	0.550 (0.117)	0.293 (0.016)	0.064 (0.011)
	<b>moy</b>	<b>0.686</b>	<b>0.291</b>	<b>0.107</b>

La fusion **active** opérée par le MFI améliore d'un facteur 2 à 9 les performances d'estimation, en particulier en présence d'experts avec de forts taux d'erreurs.

# Perception sonore active en robotique : bilan

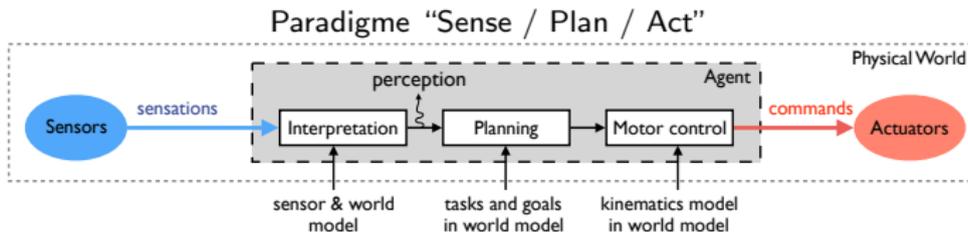


## Contributions :

- focalisées sur l'audition **binaurale**,
- en contexte audio **réaliste**,
- portent sur l'étude des **caractéristiques binaurales**, les méthodologies pour la **localisation** et l'**analyse de scène multimodale**,
- et s'appuient, à des degrés divers, sur l'**action du robot**.

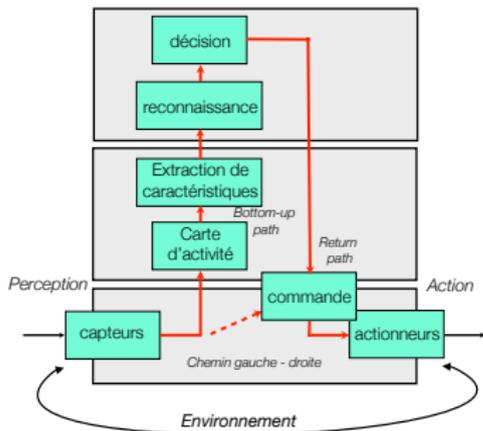
# Perception sonore active en robotique : bilan

→ Mais au final, approche traditionnelle de la perception en robotique :



- **Interprétation** des sensations sur la base de modèles connus a priori : loi de propagation du son, modélisation des effets de salle, connaissance des HRTFs, etc.
- **Plannification** du mouvement sur la bases de relations géométriques connues a priori : forme de l'environnement, lien position du robot / position de la source identifié, etc.
- **Contrôle** du mouvement effectué grâce à la connaissance préalable de la structure mécanique du robot, des modèles d'actionneurs, etc.

# Perception en robotique

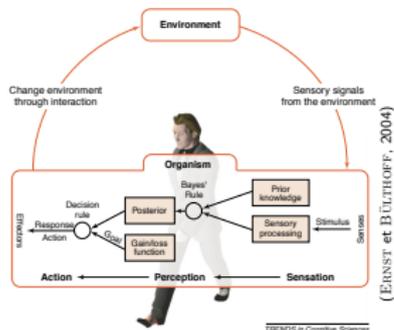


*"An agent is an active perceiver if it **knows why** it wishes to sense, and then **chooses what** to perceive, and determines **how, when and where** to achieve that perception."* (BAJCSY, ALOIMONOS et TSOTSOS, 2018)

- premières bases posées dans les années 90 (CHAUMETTE, 1998)
- permet de traiter des données sensorielles bruitées, ambiguës, etc ...
- ... via l'exploitation de modèles, parfois complexes,
- ... mais reste (relativement) peu générique.

*Approche bayésienne :*

- permet de traiter l'incomplétude des modèles,
- et prend en compte des connaissances a priori.



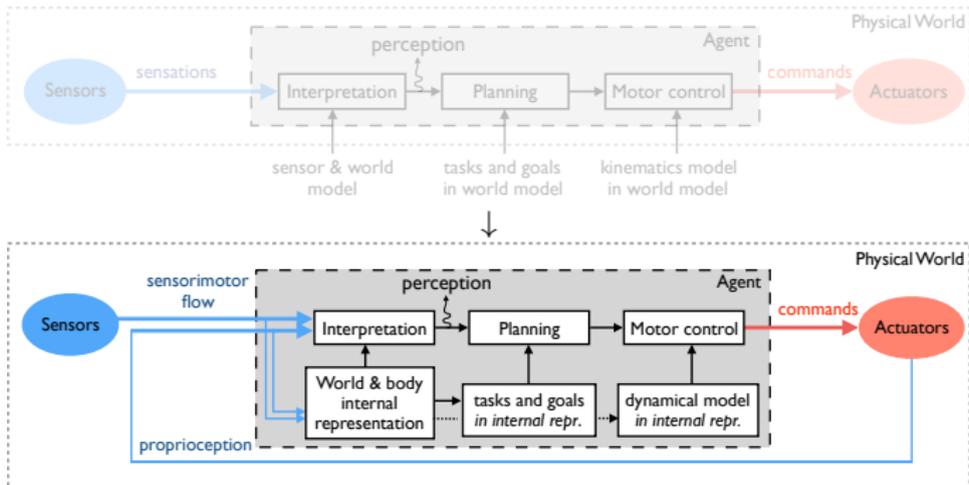
(ERNST et BÜLTHOFF, 2004)

TRENDS in Cognitive Sciences

- Le robot ne perçoit pas, l'ingénieur le fait pour lui
- Perception et action sont toujours découplées, l'une "commandant" l'autre

→ **Action** = moyen pour améliorer la perception, moyen d'acquérir des connaissances pour réaliser une tâche.

# Le point de vue de l'agent



La perception n'est plus une capacité passive donnée a priori par l'ingénieur, mais une capacité construite, découverte et limitée par les capacité d'action et de sensation de l'agent dans son environnement. L'agent est alors :

- incarné : il possède un corps
- situé : il est acteur dans son environnement

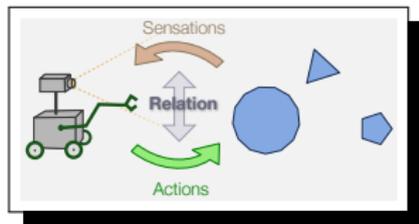
Perception interactive (BOURG et al., 2017)

# La théorie des contingences sensorimotrices

Percevoir, c'est entrer dans une **interaction sensorimotrice**, de laquelle on peut extraire des **contingences**.



(O'REGAN, 2011)

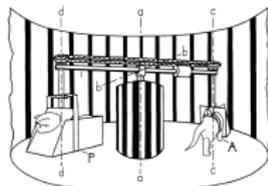


Ces *relations stables* qui lient événement moteur / événement sensoriel deviennent les éléments structurant de la perception.

→ tout est présent dans le flux sensorimoteur



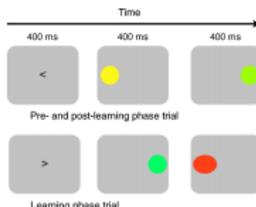
(BLAKEMORE et F. COOPER, 1970)



(HELD et HEIN, 1963)



(BACH-Y-RITA, 1975)



(RICHTERS et ESKEW, 2009)

# Question (comment formaliser tout cela ?)

Quels sont les éléments structurant le flux sensorimoteur d'un agent naïf qui permettent d'expliquer comment émerge la notion de perception ?

Éléments de réponse :

- Estimation de la dimension de l'espace
- Structure des invariants sensorimoteurs : approche intuitive

→ Thèse d'Alban LAFLAQUIÈRE (09/2009–06/2013)

- Structure des invariants sensorimoteurs : approche formelle

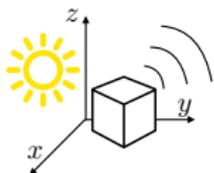
→ Thèse de Valentin MARCEL (09/2015–auj.)

Hypothèses :

- agent naïf : aucune connaissance a priori sur son corps, ses capteurs, ses actionneurs
- proprioception séparée de l'extéroception
- aucune tâche préalablement établie

# Un problème préalable : l'espace

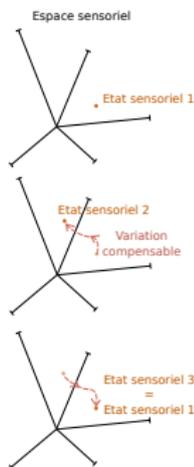
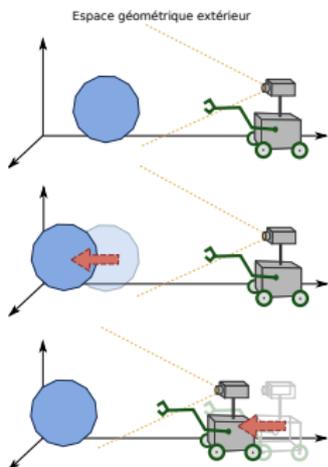
→ Point de vue traditionnel :



Contenant pré-existant dans lequel est immergé l'agent et son environnement.

Défini sur la base d'un référentiel, au sein duquel les notions de distance, position, etc. sont naturellement posées.

→ Point de vue incarné et sensorimoteur :

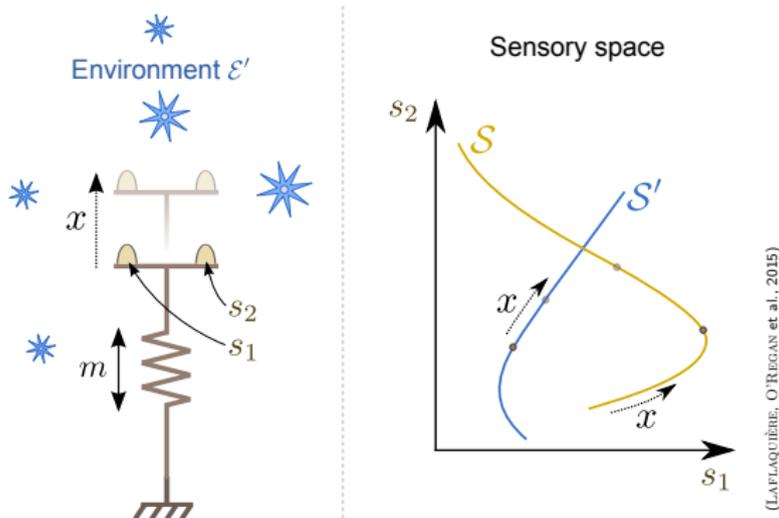


Ensemble des **propriétés** partagées par l'agent et son environnement (le flux sensorimoteur n'est pas homogène, isotrope et 3D).

- Transformations compensables : (POINCARÉ, 1887 ; POINCARÉ, 1902)
- Importance des objets rigides
- Notion de groupe : transformation nulle, combinaison des transformations, etc.
- Première formalisation : (PHILIPONA, 2008)
- Extension aux grands mouvements : (LAFLAQUIÈRE, ARGENTIERI et al., 2012)

# Et après ? Vers une structure des invariants sensorimoteurs (1/2)

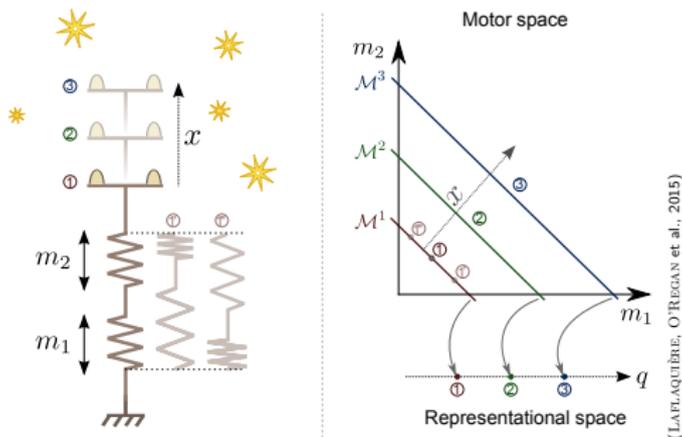
L'approche précédente s'appuie exclusivement sur la variété sensorielle. Mais ...



... la "représentation" interne de  $x$  (variable spatiale qui caractérise l'interaction de l'agent dans son environnement) dépend de la configuration  $\epsilon$  de l'environnement.

# Et après ? Vers une structure des invariants sensorimoteurs (2/2)

Alors que si on raisonne dans l'espace des configurations motrices  $\mathcal{M}$ , il vient :



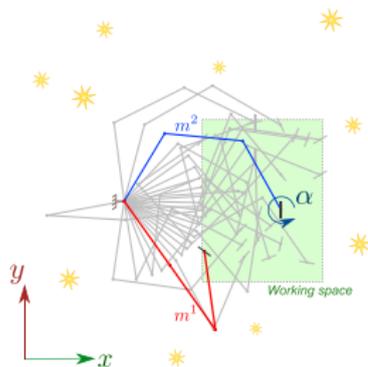
- $m_1 + m_2 = x_{\text{①}}$  : droites paramétrées par  $x_{\text{①}}$  dans  $\mathcal{M}$  ;
- ces droites constituent plus généralement des ensembles noyaux ;
- si on définit une distance entre courbes noyaux, on peut "capturer" la variable spatiale  $x$ .

## Hypothèses :

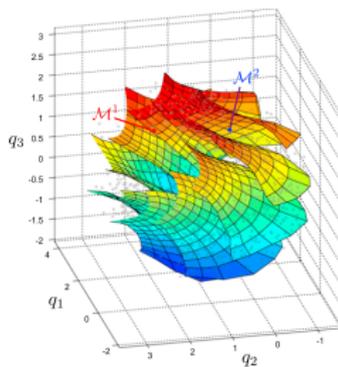
- environnement suffisamment riche
- 2 ens. noyau proches  $\Leftrightarrow$  2 config. sensorielles proches  $\Leftrightarrow$  2 config. spatiales proches

# Résultats préliminaires

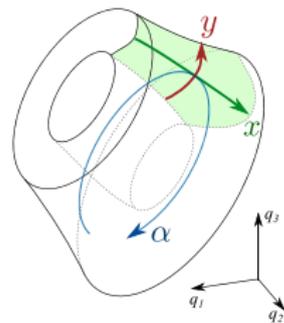
Point de vue externe



Point de vue interne



Interprétation



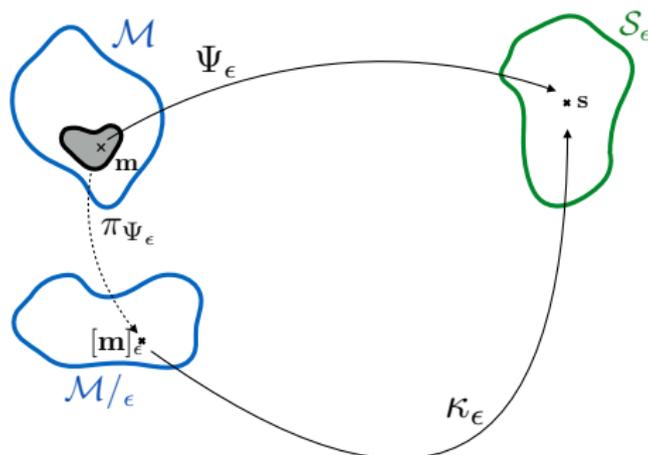
(LAFLAQUIÈRE, O'REGAN et al., 2015)

Ça marche ! Mais ...

- comment échantillonner les ensembles noyaux ?
- environnement supposé fixe et suffisamment informatif
- peu de formalisation mathématique : que représente-t'on ?

# Vers un formalisme sensorimoteur

Le point de vue “interne” :



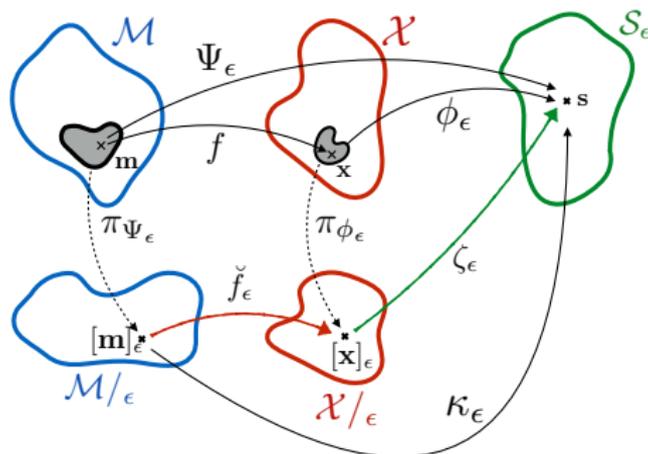
Définition d'un brique logique “égalité des sensations”  $\rightarrow$  relation d'équivalence :  
 $m_1 =_{\Psi_\epsilon} m_2 \Leftrightarrow \Psi_\epsilon(m_1) = \Psi_\epsilon(m_2)$ .

Ce qui permet la définition d'un ensemble quotient :  
 $\mathcal{M}/\epsilon = \{[m]_\epsilon \in \mathcal{P}(\mathcal{M}) \mid m \in \mathcal{M}\}$ , avec  $[m]_\epsilon = \{r \in \mathcal{M} \mid r =_{\Psi_\epsilon} m\}$ .

L'ensemble quotient  $\mathcal{M}/\epsilon$  est ce qui est accessible à l'agent, découvert via ses invariants sensorimoteurs  $\rightarrow$  c'est ce qui a été construit intuitivement dans (LAFLAQUIÈRE, O'REGAN et al., 2015).

# Vers un formalisme sensorimoteur

Le point de vue "externe" :  $\Psi_\epsilon = \phi_\epsilon \circ f$



On introduit "l'espace des poses"  $\mathcal{X}$ , où on peut définir une nouvelle relation d'équivalence :  $\mathbf{x}_1 =_{\phi_\epsilon} \mathbf{x}_2 \Leftrightarrow \phi_\epsilon(\mathbf{x}_1) = \phi_\epsilon(\mathbf{x}_2)$ .

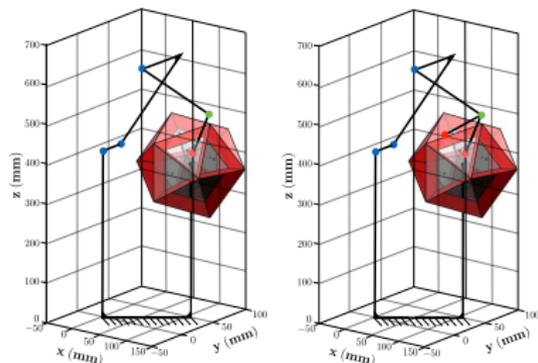
Qui a son tour permet la définition d'un nouvel ensemble quotient :  $\mathcal{X}/\epsilon = \{[\mathbf{x}]_\epsilon \in \mathcal{P}(\mathcal{X}) \mid \mathbf{x} \in \mathcal{X}\}$ , avec  $[\mathbf{x}]_\epsilon = \{\mathbf{r} \in \mathcal{X} \mid \mathbf{r} =_{\phi_\epsilon} \mathbf{x}\}$ .

Sous certaines hypothèses topologiques (compacité, séparabilité, etc.), on peut montrer que :  $\mathcal{M}/\epsilon$  est homéomorphe à  $\mathcal{X}/\epsilon$

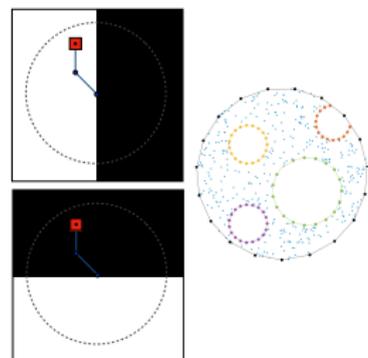
# Deux applications à ce formalisme

Représentation de l'interaction d'un agent avec ...

... son propre corps

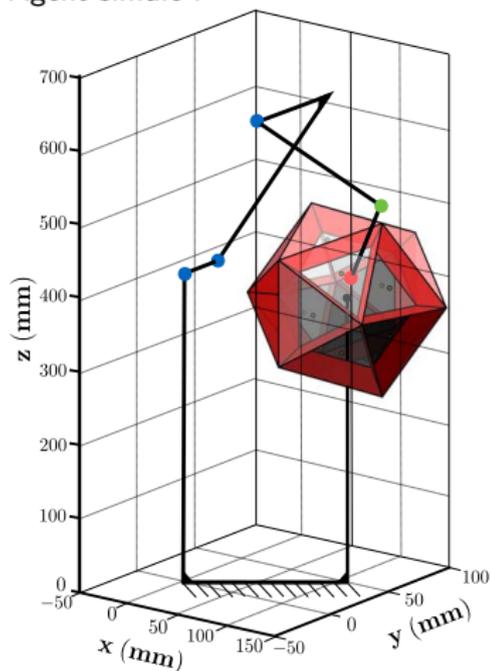


... son espace de travail



# Application à la représentation du corps

Agent simulé :

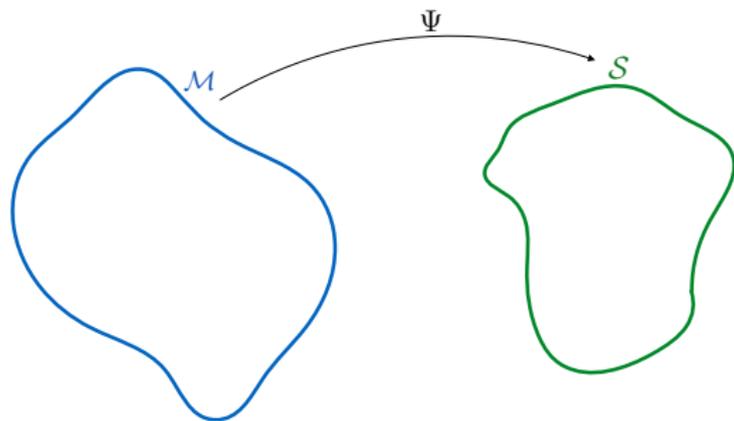


Paramètres :

- 6 moteurs, corps "sphérique"
- exploration motrice : marche aléatoire
- environ  $10^6$  configurations motrices
- sensation si contact (ponctuel) avec le corps
- $s$  de dimension 20

# Application à la représentation du corps

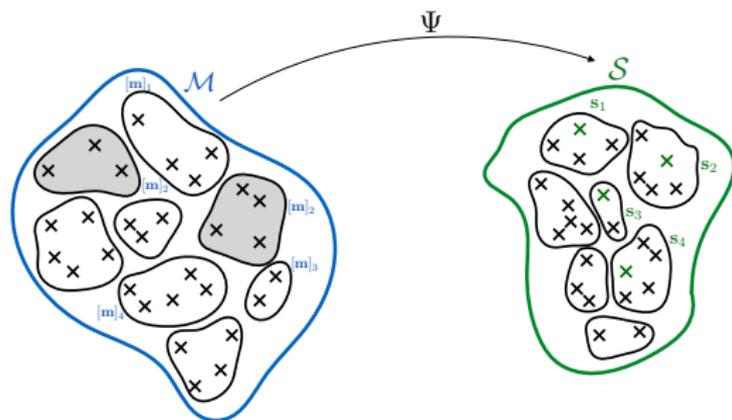
Construction de la représentation :



L'agent n'a accès qu'à ses commandes motrices  $\mathbf{m}$  et sensations associées  $\mathbf{s}$

# Application à la représentation du corps

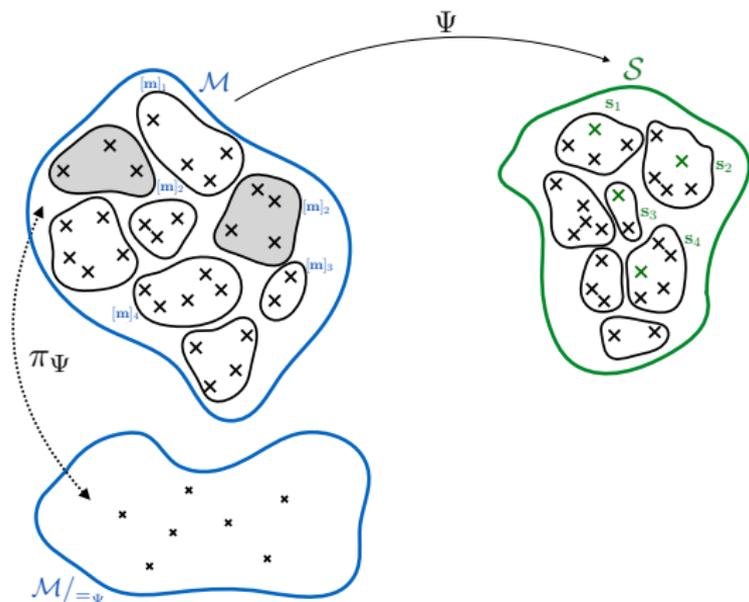
Construction de la représentation :



- On explore aléatoirement l'espace moteur ... et donc l'espace sensoriel ;
- Une mesure de "similarité sensorielle" permet de regrouper les sensations ;
- On définit ainsi les classes d'équivalence motrices  $[m]_j$ .

# Application à la représentation du corps

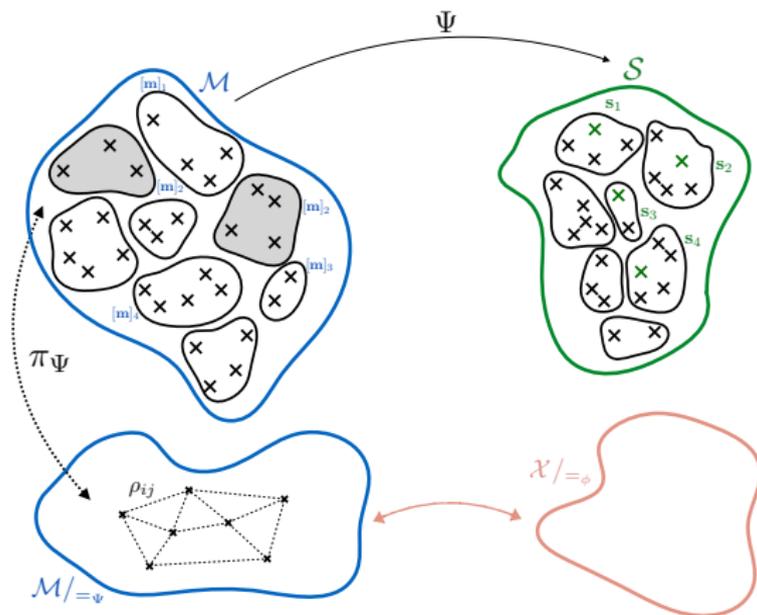
Construction de la représentation :



Les classes d'équivalences  $[m]_i$  sont en fait les éléments de l'espace quotient  $\mathcal{M}/_{=\Psi}$

# Application à la représentation du corps

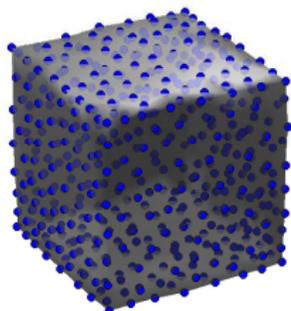
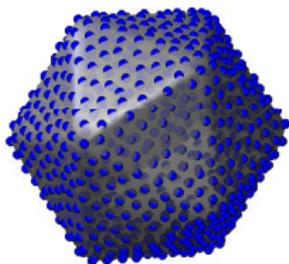
Construction de la représentation :



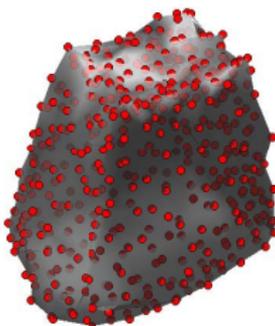
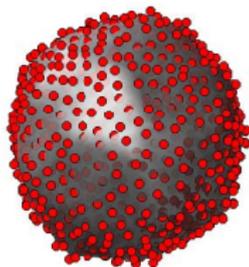
- La distance quotient  $\rho$  permet d'approximer  $\mathcal{M}/_{=_{\psi}}$  sous la forme d'un graphe ;
- Ce graphe capture l'information topologique de  $\mathcal{X}'/_{=_{\phi}} \rightarrow$  **visualisation via CCA.**

# Application à la représentation du corps : résultats

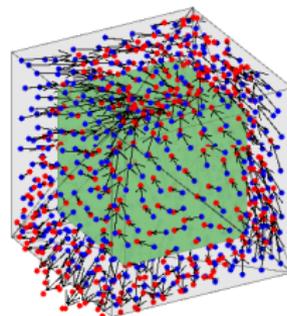
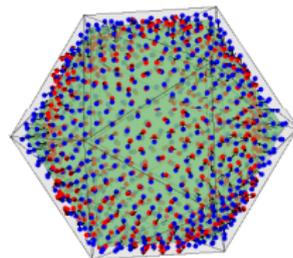
Corps de l'agent



Représentation (CCA)



Comparaison



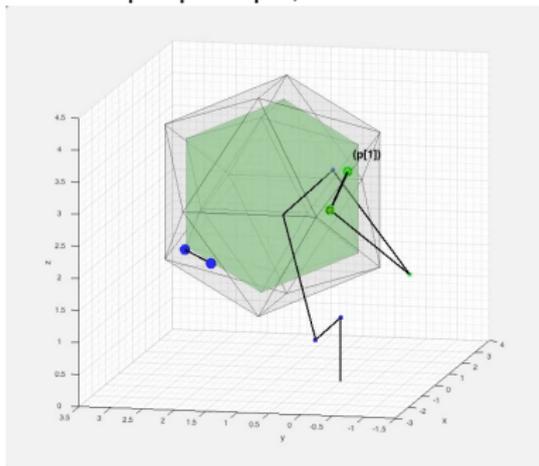
(MARCEL, ARGENTIERI et GAS, 2017)

→ Attention : cas “simple” car contact ponctuel. Si l'interaction est plus complexe, la représentation l'est également (du moins dans son interprétation)...

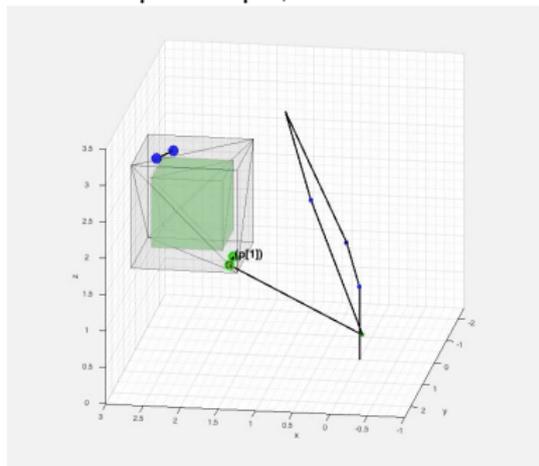
# Exploitation de la représentation : interpolation motrice

Résultats :

Corps sphérique, 2 contacts :



Corps cubique, 2 contacts :



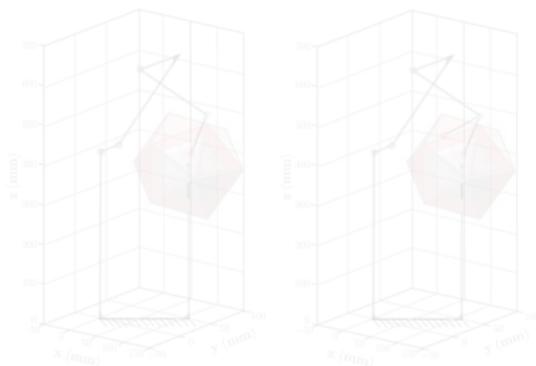
(MARCEL, ARGENTIERI et GAS, 2017)

- continuité du mouvement sur le corps garantie par l'homéomorphisme (du moins, dans le cas "continu")
- stratégie utilisable même pour des interactions plus complexe
- stratégie utilisable pour une autre modalité que tactile : exploitation avec la vision validée ("voir = toucher avec les yeux") ... sous certaines hypothèses

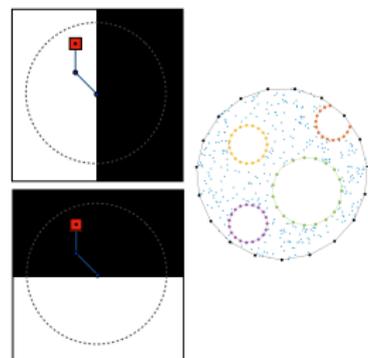
# Deux applications à ce formalisme

Représentation de l'interaction d'un agent avec ...

... son propre corps

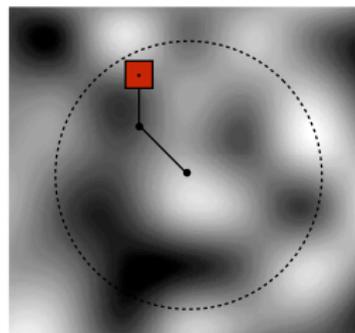
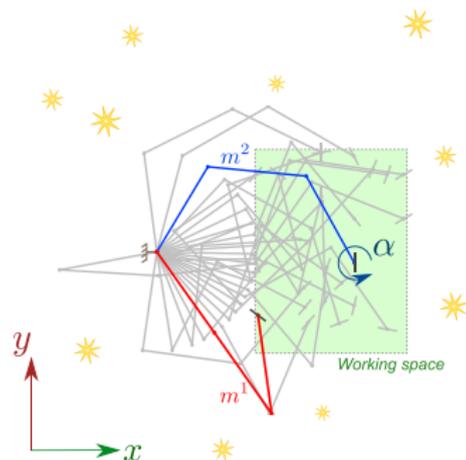


... son espace de travail



# Application à la représentation de l'espace de travail

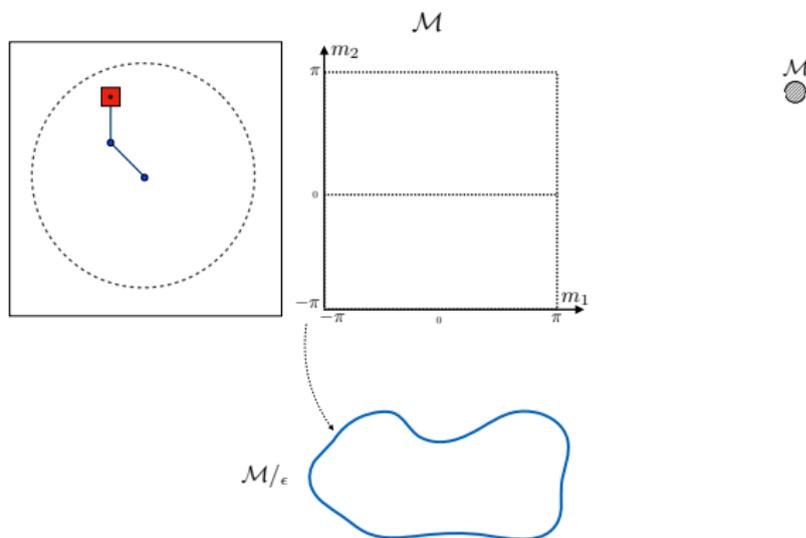
Retour sur le contexte, analogue à l'illustration précédent :



- Moteur : bras à 2 degrés de liberté, de sorte que  $\mathbf{m} \in ]-\pi; \pi]^2$
- Environnement : succession d'images N&B aléatoires
- Sensation :  $s = 1$  si le niveau de gris est supérieur à 0.5, et donc  $\mathbf{s} \in \{0, 1\}$

# Application à la représentation de l'espace de travail

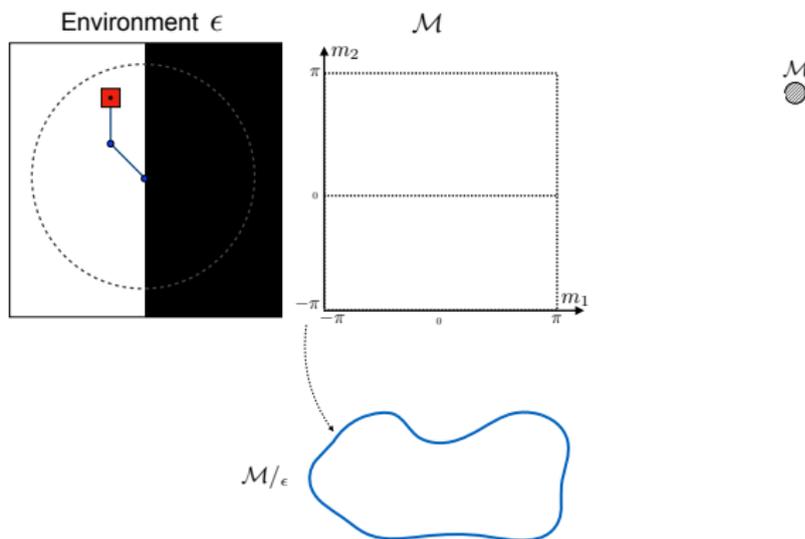
Principe :



Au départ, l'agent est naïf : son espace moteur  $\mathcal{M}$  n'est pas structuré.

# Application à la représentation de l'espace de travail

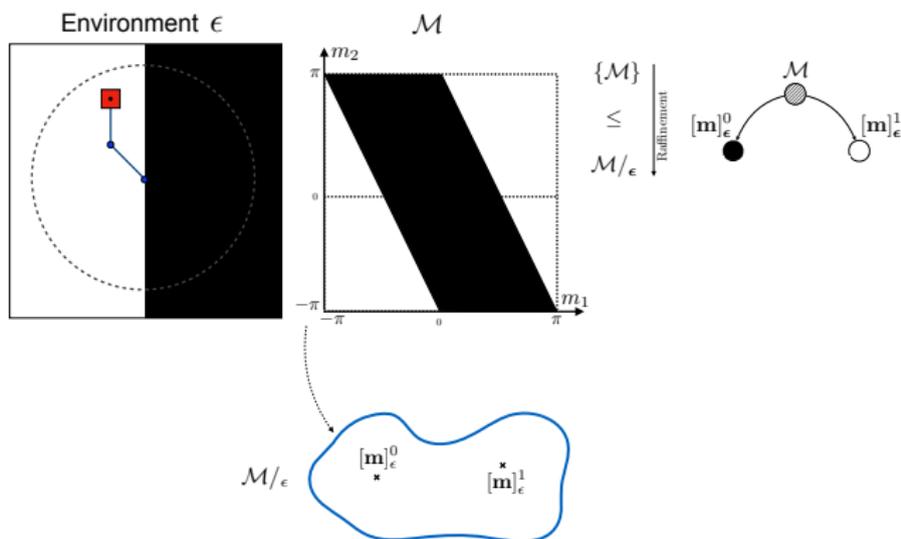
Principe :



L'environnement est dans une configuration  $\epsilon$  et sépare l'espace de travail en deux parties distinctes (point de vue externe).

# Application à la représentation de l'espace de travail

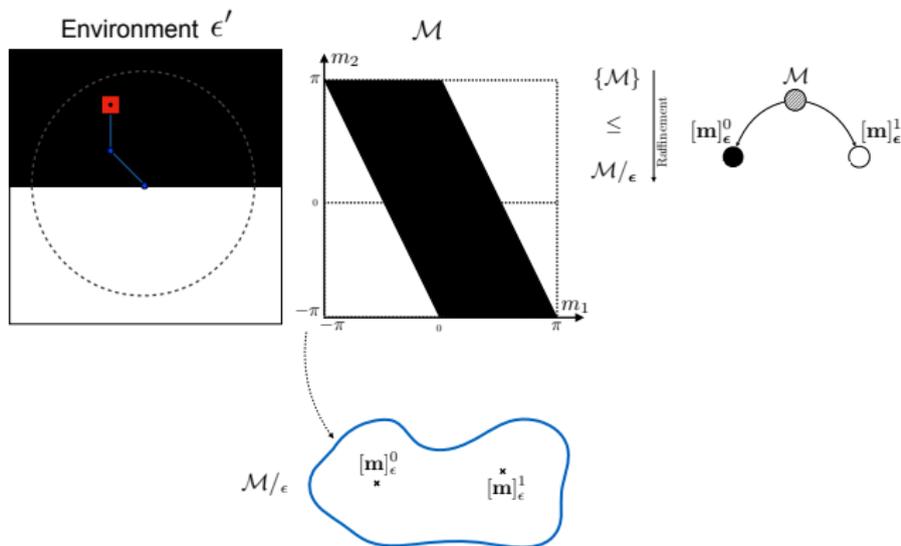
Principe :



Après exploration motrice, l'agent partitionne son espace moteur en 2 classes d'équivalence  $[\mathbf{m}]_\epsilon^0$  et  $[\mathbf{m}]_\epsilon^1$ , associées aux sensations  $\mathbf{s} = 0$  ou  $1$ .

# Application à la représentation de l'espace de travail

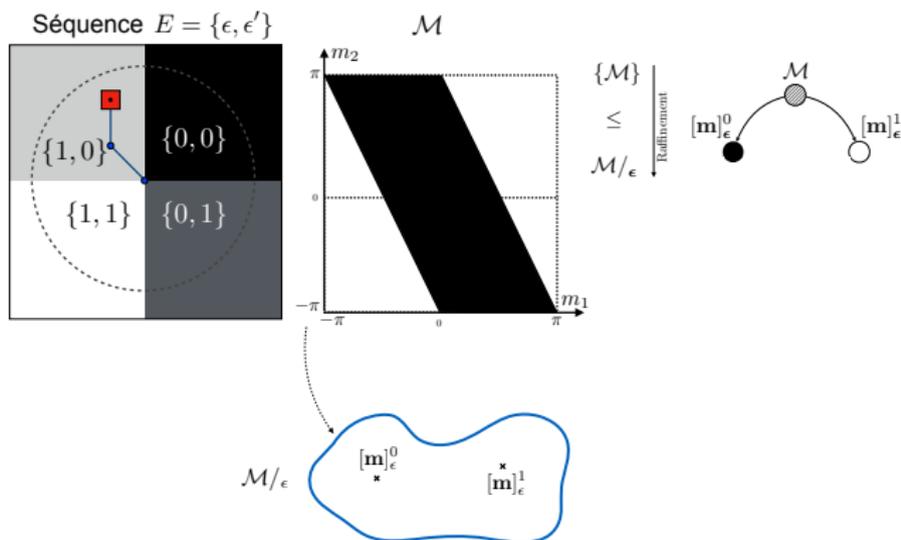
Principe :



L'environnement change de configuration  $\epsilon \rightarrow \epsilon'$

# Application à la représentation de l'espace de travail

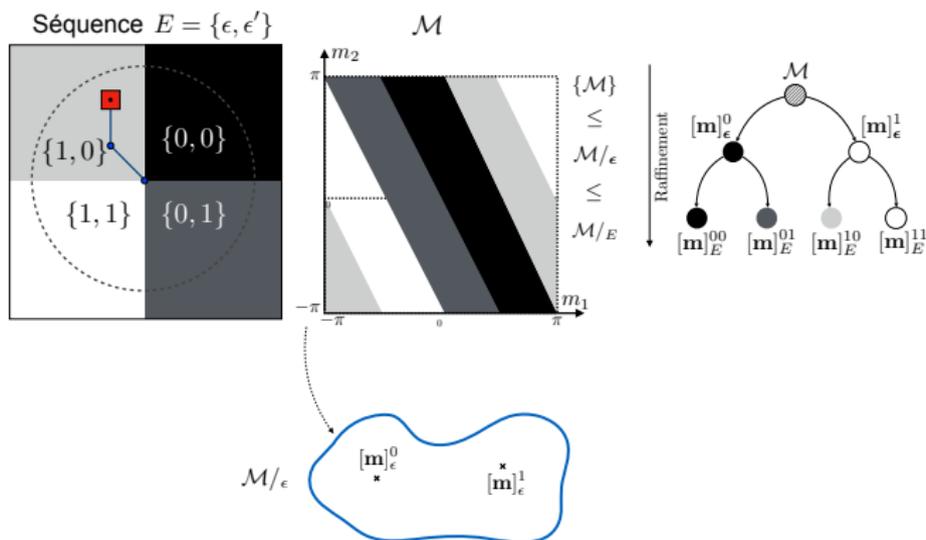
Principe :



L'agent a fait l'expérience de la séquence  $E = \{\epsilon, \epsilon'\}$ , qui sépare en 4 parties l'espace de travail (point de vue externe).

# Application à la représentation de l'espace de travail

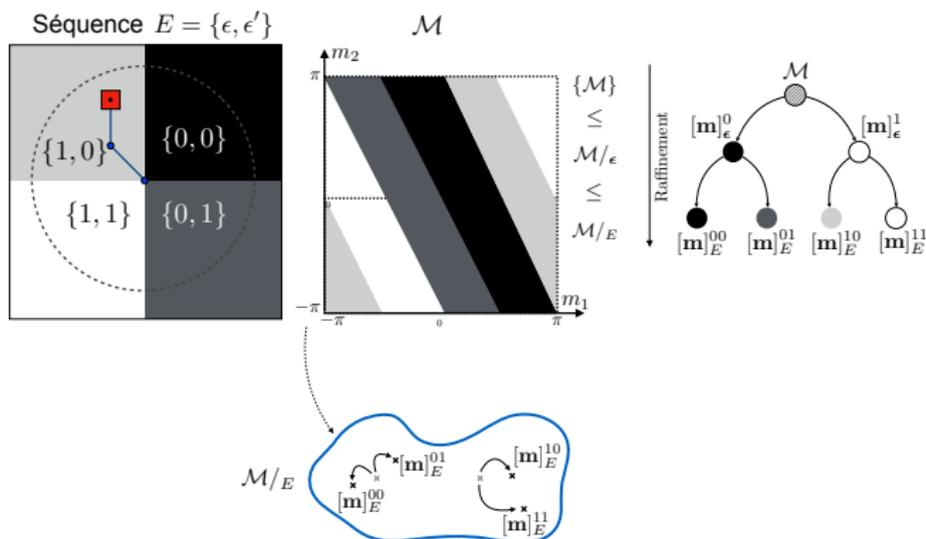
Principe :



A cette séquence correspond en fait un raffinement de la partition motrice obtenue en ayant interagi avec l'environnement dans sa configuration  $\epsilon$ .

# Application à la représentation de l'espace de travail

Principe :

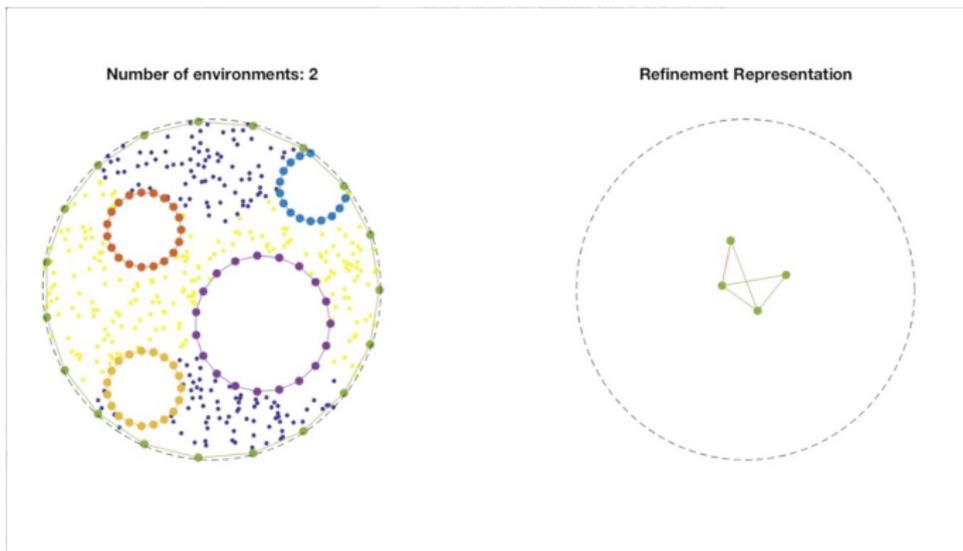


Cela se traduit, dans l'espace quotient, par un fractionnement des classes d'équivalence : des configurations motrices **indiscernables** par l'expérience sensorimotrice précédente deviennent **distinctes** dans la représentation courante.

La représentation  $\mathcal{M}/E$  évolue tout au long de l'interaction de l'agent avec son environnement

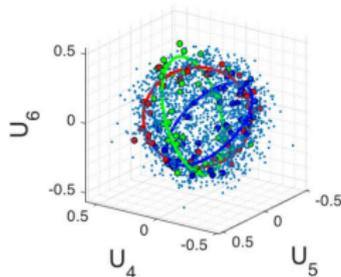
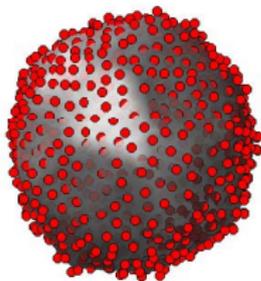
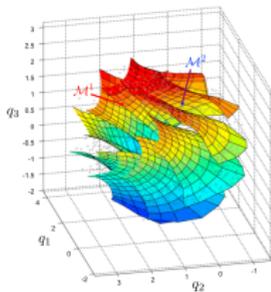
# Représentation de l'espace de travail : résultats

Si on introduit une métrique statistique dans  $\mathcal{M}/E$  (on compte simplement le nombre de fois où les sensations sont égales), on peut alors représenter le graphe de la représentation.



On converge vers une représentation topologiquement correcte de l'espace de travail, qui est certainement *la plus fine* accessible à l'agent.

# Perception sensorimotrice : bilan



## Contributions :

- ancrées au sein d'un paradigme sensorimoteur **bas-niveau** ;
- visent à comprendre la structure des **invariants sensorimoteurs** ;
- reprise des travaux de Philipona sur l'**estimation de la dimension de l'espace** ;
- vers un **formalisme** (topologique) **de l'expérience sensorimotrice** ;
- **applications** à la découverte du corps et de l'espace opérationnel d'un agent ...
- ... de type "bras" seulement.

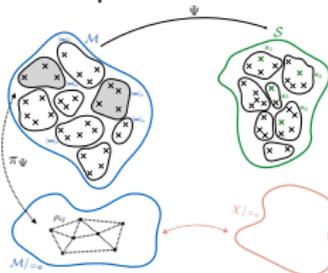
# Conclusion

Perception active binaurale



← Action →

Perception interactive



- l'audition binaurale ne peut plus être qu'active
- exploiter uniquement 2 microphones reste un challenge
- qui veut un système audio robotique efficace doit exploiter des antennes de microphones

- approche fondamentale et bas niveau ...
- ... dont la robustesse aux conditions expérimentales réalistes doit être (é)prouvée,
- et généralisée à des systèmes capables de déplacements "absolus" dans l'environnement

Et pourquoi ne pas rapprocher ces 2 champs de recherche ?

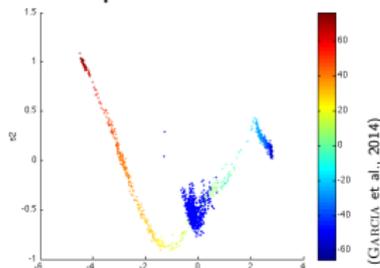
# Conclusion

Perception active binaurale



← Action →

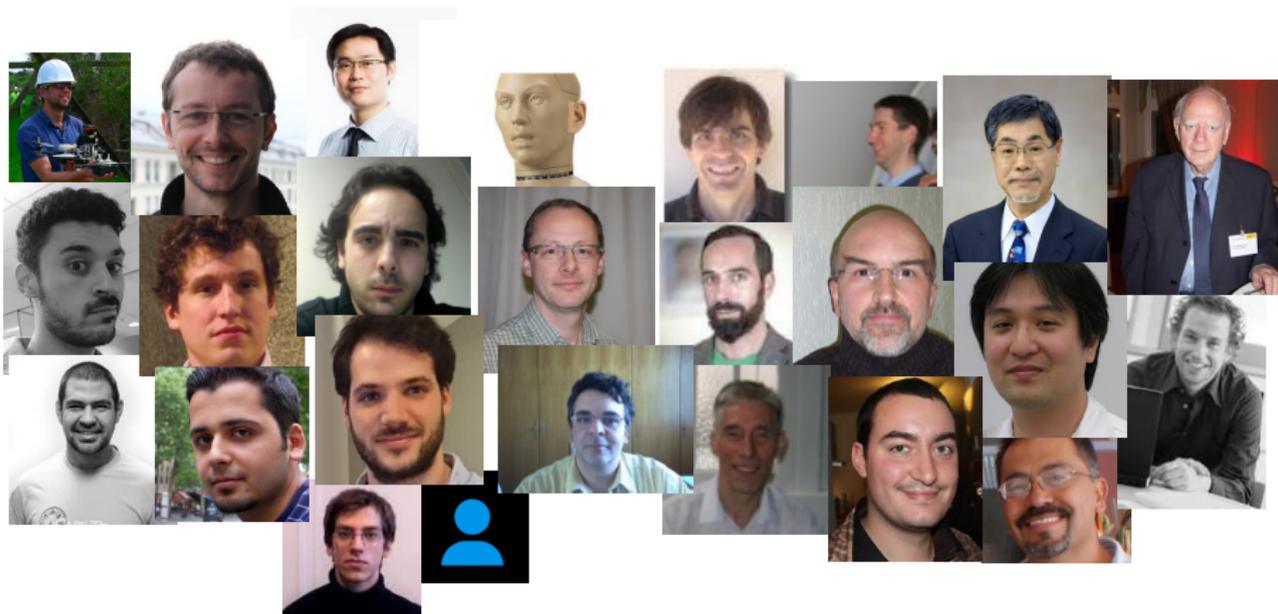
Perception interactive



- l'audition binaurale ne peut plus être qu'active
- exploiter uniquement 2 microphones reste un challenge
- qui veut un système audio robotique efficace doit exploiter des antennes de microphones
- approche fondamentale et bas niveau ...
- ... dont la robustesse aux conditions expérimentales réalistes doit être (é)prouvée,
- et généralisée à des systèmes capables de déplacements "absolus" dans l'environnement

Et pourquoi ne pas rapprocher ces 2 champs de recherche ?

# Tout cela est avant tout un travail d'équipe !



## Questions ?