

Reconnaissance Interactiviste de Rythmes Musicaux

Interactivist Recognition of Musical Rhythms

Jean-Christophe BUISSON^{1,2}

Jean-Luc BASILLE^{1,2}

¹ ENSEEIHT, Toulouse

² IRIT, Toulouse

ENSEEIHT, 2 rue Camichel, 31071 TOULOUSE

buisson@enseeiht.fr, basille@enseeiht.fr

Résumé

Cet article présente les idées du modèle interactiviste de Mark Bickhard et du cadre théorique assimilation/accommodation de Jean Piaget au travers d'un programme de reconnaissance de rythmes musicaux frappés au clavier d'un ordinateur. De façon totalement non supervisée, le programme apprend progressivement à reconnaître des rythmes de plus en plus complexes. Il le fait sans jamais enregistrer le flot d'entrée, et sans opération de pattern-matching au sens usuel. Au contraire, des processus internes sont construits dynamiquement pour suivre et anticiper les actions de l'utilisateur. Nous montrons dans cet article que de tels processus sont des représentations des rythmes au sens 'interactiviste' du terme, et qu'ils émergent de sources non représentatives, évitant ainsi le 'symbol grounding problem'. Ce ne sont pas des copies ou des transductions de la réalité, mais des constructions idéales de l'agent, évitant également une circularité décrite par Piaget. D'un point de vue pratique, la nature active de ce processus de reconnaissance lui permet de fonctionner en présence de flots d'entrée complexes et/ou bruités.

Abstract

This paper advocates the main ideas of the interactive model of representation of Mark Bickhard and the assimilation/accommodation framework of Jean Piaget, through a rhythm recognition demonstration software. In a completely unsupervised way, the program progressively learns to recognize more and more complex rhythms struck on the user's keyboard. It does so without any recording of the input flow, and without any pattern matching in the usual sense. On the contrary, internal processes are dynamically constructed to follow and anticipate the user's actions. We show that these processes are representations of the rhythms in the interactivist sense, and that they

emerge from non representational grounds, avoiding the symbol-grounding problem. They are not copies or transductions of reality, but ideal internal constructions of the agent, avoiding a circularity pointed out by Piaget. Practically, the active nature of this recognition process allows it to work even with noisy and complex input flows.

Keywords

reconnaissance de rythme, schémas sensori-moteurs, Interactivisme, Piaget

1 Introduction

Selon le modèle interactiviste [2, 3], les représentations cognitives ne peuvent être composées simplement d'encodages symboliques, car alors elles manqueraient de tout contact véritable avec la réalité. Bickhard a montré qu'une forme plus élémentaire de représentation était nécessaire, interactive et temporelle par nature. Piaget [15] a également montré que les représentations mentales ne peuvent pas être des copies simplifiées du monde, mais sa théorie de la perception [16] est également tombée dans le piège des encodages et de leurs contradictions. Néanmoins, son cadre théorique des schémas assimilateurs, interactifs et temporels par nature, a inspiré notre propre modèle. Notre programme de démonstration a été conçu pour illustrer ces idées, et non pour servir de véritable logiciel de reconnaissance de rythmes musicaux [21, 6]. Il nous a obligés à fournir un modèle détaillé et opérationnel de représentation interactive, ajoutant la quantification à des théories qualitatives. Le programme interagit avec des utilisateurs humains, comme cela est recommandé par Rodney Brooks, de façon à garantir qu'aucun aspect important n'a été oublié, ou modélisé incorrectement: "The physical grounding of a system in the world forces its designer to deal with all the issues. ... There is no room for

cheating” [4].

Dans la première partie de cet article, nous décrivons en détails le programme de démonstration, ses performances et le modèle de perception associé. Dans une seconde partie, nous examinons d’importants problèmes de la psychologie de la perception, et nous utilisons nos résultats pour plaider en faveur du cadre interactiviste.

2 Le programme de reconnaissance de rythmes

2.1 Présentation

Le programme de démonstration apprend à reconnaître des rythmes musicaux, de façon complètement non supervisée. La figure 1 montre une copie d’écran du programme, lors de la reconnaissance d’un rythme simple.

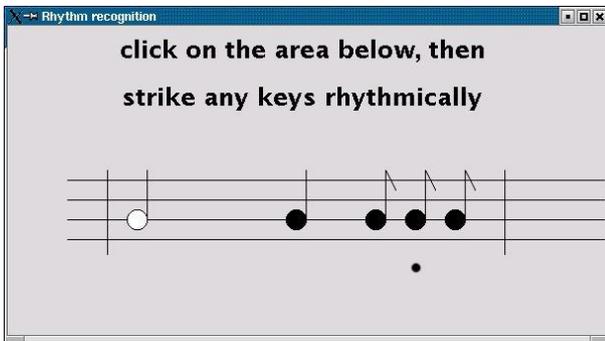


Figure 1: Copie d’écran du programme de reconnaissance de rythmes. Le point noir sous les notes anticipe les frappes de l’utilisateur en temps réel.

L’utilisateur interagit avec le programme en frappant une seule touche au clavier. Lorsqu’un rythme est reconnu, une partition avec une seule mesure est affichée à l’écran, et un point noir se déplace de note en note, anticipant les frappes de l’utilisateur. Celui-ci ne donne aucune indication d’aucune sorte, et l’apprentissage s’effectue donc de façon totalement non supervisée. Le programme ne reconnaît pas tout de suite des rythmes complexes; il doit d’abord être exposé à des rythmes simples. Lorsque des rythmes simples ont été reconnus, l’utilisateur peut essayer des rythmes de plus en plus complexes, ressemblants ou basés sur les précédents. Plus la culture musicale du programme est riche, et plus celui-ci est capable d’assimiler des rythmes nouveaux et complexes.

Du bruit peut être ajouté sous forme de frappes aléatoires, et l’apprentissage continue de fonctionner lorsque le niveau de ce bruit ne dépasse pas 20% de l’ensemble des frappes.

Le programme a été conçu pour discriminer les notes suivantes:

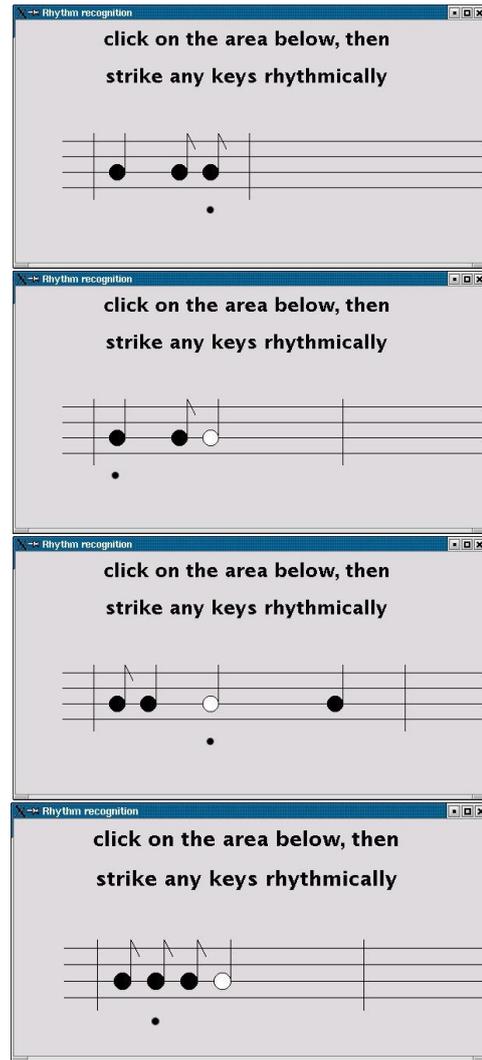


Figure 2: Autres copies d’écran du programme de reconnaissance de rythmes.

1. blanches
2. noires
3. croches

Il faut environ 4 à 5 périodes pour qu’un nouveau rythme soit assimilé et affiché par le programme, à la condition qu’il soit effectivement assimilable, c’est à dire qu’il ressemble à des rythmes préalablement assimilés. L’utilisateur n’a pas besoin d’avoir un timing parfait, bien que les meilleurs tuteurs soient ceux qui ont une formation musicale.

Lorsqu’un rythme qui a été assimilé dans le passé est joué à nouveau, le programme met environ une à deux périodes pour se synchroniser sur lui. On peut donc dire que le rythme avait été mémorisé, et qu’il est maintenant rémémoré.

Le point noir sous les notes anticipe les frappes de l'utilisateur. Si l'utilisateur s'arrête brusquement de frapper les touches, le point se déplace d'une ou deux notes avant de disparaître avec la portée. Un phénomène analogue se produit lorsque, pour un rythme bien assimilé, l'utilisateur 'oublie' de frapper une note, et que le point noir se déplace sous la note absente, comme si rien ne s'était passé. Nous nous comportons de la même façon lorsque nous lisons sans remarquer les erreurs typographiques.

Lorsque le tempo de l'utilisateur varie légèrement, le programme s'accommode à ce changement, et continue d'être en synchronie. Ces variations de tempo ne doivent pas être trop brutales. Bien sûr, une synchronisation mutuelle se produit souvent: pendant que le programme s'accommode aux variations de tempo de l'utilisateur, l'utilisateur tend inconsciemment à se synchroniser sur le programme.

2.2 Fonctionnement du programme

Comment il ne fonctionne pas. Le programme n'effectue aucun appariement des frappes en entrée avec des motifs préstockés. En fait, il n'effectue même aucun enregistrement des frappes en entrée.

Schémas sensori-moteurs. Le programme gère un ensemble de Schémas Sensori-Moteurs (SSM), dont les éléments évoluent dynamiquement selon un algorithme décrit à la section suivante. A chacun de ces SSMs est associée une partition musicale pour une mesure.

D'un point de vue informatique, un SSM est un thread, qui joue sa partition de façon silencieuse, en temps réel et avec un certain tempo, en essayant d'être synchronisé avec les frappes en entrée de l'utilisateur.

Supposons que la partition d'un SSM corresponde au rythme qui est actuellement joué par l'utilisateur, ET que leurs tempos soient identiques, ET que la position courante dans la partition du SSM corresponde à la position dans le flot des frappes de l'utilisateur. Cette situation est illustrée figure 3.

Dans cette situation, le SSM est parfaitement synchronisé avec la réalité - ce qui est sa raison d'être - mais toute variation dans le tempo des frappes peut conduire à une désynchronisation. Dans la situation de synchronisation, le SSM assimile la réalité selon le terme consacré par Piaget, mais il a besoin d'un mécanisme d'accommodation pour préserver ce processus d'assimilation.

Nous avons incorporé un tel *mécanisme d'accommodation temporel*, qui permet à un SSM presque synchronisé avec les frappes en entrée, de se retrouver en parfaite synchronie. Plus précisément, lorsque qu'une frappe de l'utilisateur arrive avant le moment anticipé par un SSM, son tempo est légèrement augmenté, et la position courante dans la partition du SSM est réalignée sur la frappe. Un

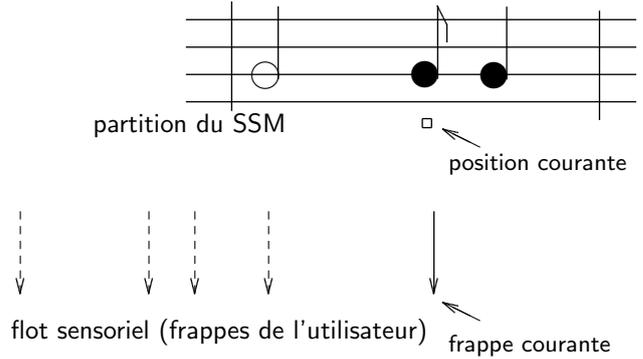


Figure 3: Schéma sensori-moteur synchronisé sur le flot des frappes de l'utilisateur

ajustement analogue inverse est effectué lorsqu'une frappe anticipée arrive après.

Maintenant que nous disposons d'un mécanisme de *préservation* de la synchronie, nous avons besoin d'un autre qui nous permette d'*amorcer* la synchronisation. Nous l'avons appelé *mécanisme de resynchronisation*: lorsqu'un SSM n'est pas du tout en phase avec les frappes de l'utilisateur, son tempo est progressivement accéléré (jusqu'à une certaine limite), et il est capable de se resynchroniser en peu de cycles. (figure 4).

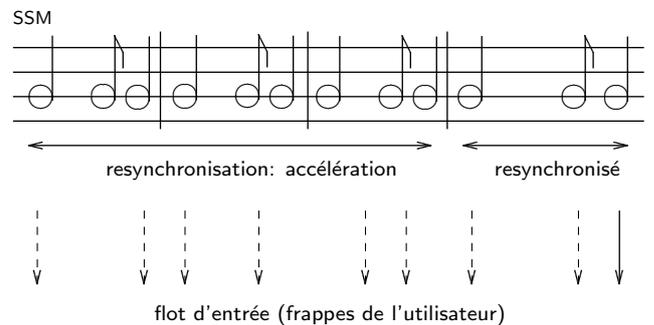


Figure 4: Resynchronisation: après quelques cycles d'un tempo accéléré, le SSM rattrape les frappes de l'utilisateur

Activité d'un schéma sensori-moteur. Un degré d'activité est mis à jour en permanence pour chaque SSM, qui évalue à quel degré il est synchronisé avec le flot des frappes en entrée. C'est un nombre entre 0 et 100, 100 signifiant une parfaite adéquation et 0 aucune adéquation; il permet de classer l'activité de l'ensemble de SSM actifs. Le calcul de l'activité d'un SSM a été fait de façon ad-hoc, et beaucoup de réglages ont été nécessaires pour que l'ordre induit par ces valeurs corresponde à l'appréciation intuitive de cette adéquation. Il faut en effet comprendre que

dans beaucoup de situations, il est difficile de décider lequel de deux SSM est le mieux synchronisé sur le flot d'entrée.

Finalement, la partition du SSM ayant le plus haut degré d'activité est affichée à l'écran, ainsi qu'un point noir qui se déplace en permanence sous la position courante dans la partition.

Evolution de la population des schémas sensori-moteurs. Maintenant que nous avons décrit le comportement des SSMs, nous devons expliquer comment ils sont créés et détruits.

La population des SSMs évolue selon un algorithme de type variation/sélection, très similaire à l'évolution de la population des lymphocytes-B dans les systèmes immunitaires de mammifères [8]. Ces algorithmes sont activement étudiés pour leurs propriétés de mémoire associative et d'apprentissage inductif notamment [19, 8]. On pensait, jusqu'au milieu du vingtième siècle, que les organismes pathogènes laissaient une empreinte sur les lymphocytes sur lesquels ils se liaient, et que ces empreintes se dupliquaient lors du clonage des lymphocytes. La fausseté de cette théorie est intéressante, car elle ressemble à une fausse idée répandue dans les modèles de la perception, que l'esprit peut former des copies internes du monde, en utilisant des mécanismes de transduction à partir des organes des sens.

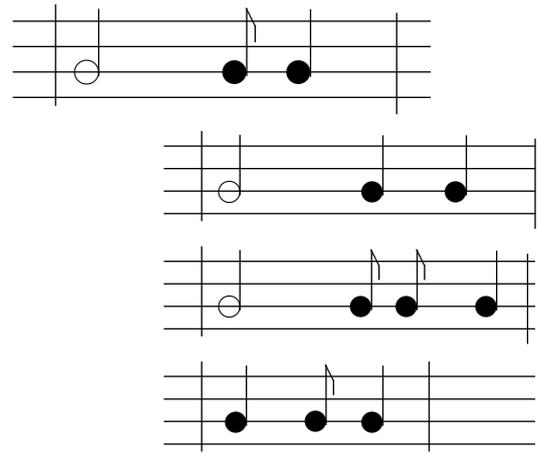
On sait maintenant que, lorsque suffisamment d'organismes pathogènes se lient à un lymphocyte-B, ce lymphocyte commence à se dupliquer rapidement avec un très fort taux de mutation. Parmi les mutants, certains vont se lier aux organismes pathogènes encore mieux que leur parent. Après un certain nombre de tels cycles, une population de lymphocytes-B finement adaptée au pathogène émerge.

Dans notre système, un SSM hautement synchronisé sur le flot sensoriel est l'analogue d'un lymphocyte-B activé par sa rencontre avec un organisme pathogène. Lorsque le niveau de synchronisation dépasse un certain seuil, le SSM commence à se dupliquer rapidement, avec un fort taux de mutation. Un SSM mutant a une partition dérivée de son parent par des opérations de substitution, suppression ou insertion (voir figure 5).

Parmi ces SSMs mutants, certains se synchroniseront avec la réalité mieux que leurs parents, tandis que d'autres se synchroniseront moins bien. Les meilleurs dépasseront rapidement le seuil et se reproduiront, et après quelques cycles, une population de SSM hautement synchronisés va émerger.

Afin de prévenir une expansion explosive de la population des SSMs, leur durée de vie est limitée, selon leur origine. Les nouveaux-nés SSMs qui se synchronisent mal avec le flot sensoriel meurent rapidement s'ils restent inactifs. Un SSM parent (un SSM qui a initié un processus de clonage) a une durée de vie

SSM parent



copies mutantes

Figure 5: Mutations possibles pour un schéma sensori-moteur.

beaucoup plus longue, car il a déjà prouvé son utilité. Ce mécanisme est encore analogue à la création de lymphocytes-B mémoire dans notre système immunitaire.

Population initiale des SSMs. Au début du processus, il n'existe qu'un seul SSM, associé au rythme à un temps. L'utilisateur doit commencer à frapper au clavier des rythmes très simples, puisqu'ils doivent être assimilés par cet unique SSM élémentaire. L'algorithme gouvernant l'évolution de la population des SSMs commence alors à produire des SSMs de plus en plus adaptés aux frappes de l'utilisateur.

Bien sûr, il est possible de fournir une population initiale de SSMs qui permette au système de reconnaître des rythmes complexes dès le début. Comme les comportements innés des mammifères, ils disparaîtront s'ils ne fonctionnent pas pendant une longue période.

Immunité au bruit. L'apprentissage et la reconnaissance fonctionnent bien lorsque du bruit est incorporé au flot d'entrée sous forme de frappes aléatoires, jusqu'à environ 5% de toutes les frappes. Cette immunité est due à la nature active du processus perceptif, qui anticipe et attend les frappes à venir, en ignorant les autres. Cependant, des frappes venant du bruit qui sont proches de frappes anticipées peuvent conduire à une désynchronisation d'un SSM; ce problème serait réduit si le flot sensoriel était plus riche (distinction des notes par exemple), car il serait plus facile de séparer une frappe venant du bruit qu'une frappe provenant de l'utilisateur.

3 Idées illustrées par le programme

3.1 Le processus perceptif est fondamentalement temporel

La perception est un processus, pas un phénomène statique. C'est très clair pour l'audition ou la perception tactile, dans lesquels la perception a nécessairement la même durée que la situation perçue, et où ses résultats dépendent des pas successifs du processus. Il ne s'agit pas seulement d'une accumulation d'informations avec le temps. Considérons par exemple une mélodie ralentie 10 fois. Pourquoi devient-elle si difficile à reconnaître? Comme Shanon l'a montré clairement [20], ce ne devrait pas être le cas, si nous avions une représentation interne statique de la mélodie. En fait, ce devrait être même plus facile !

Dans le cas de la perception tactile, Merlau-Ponty [11] a montré qu'il est nécessaire de manipuler un tissu avec un certain timing pour en déterminer correctement la nature.

La nature temporelle de la perception est peut être moins évidente pour des modalités perceptives telles que la vision, où nous avons été habitués à utiliser l'analogie de l'oeil et de la caméra, mais James Gibson et d'autres ont montré que tel n'est pas le cas [9]. Son intuition la plus importante a été peut être de montrer que nous détectons des invariances dans le temps (ainsi que dans l'espace) dans les flots perceptifs, et que ces détectations ne peuvent se faire sur quelques images immobiles. De même, beaucoup d'animaux ne réagissent qu'aux être ou aux objets en mouvement.

Cette temporalité intrinsèque de la perception est visible dans notre démonstration: chaque schéma sensori-moteur possède son tempo propre. Il peut s'adapter à des tempos un peu plus ou un peu moins rapides, mais cette adaptation n'est pas extensible indéfiniment.

3.2 Sensation et action sont inséparables dans le processus perceptif

Pour sentir une odeur, il faut inspirer. Pour palper un objet, on a besoin de déplacer les doigts sur lui, de le manipuler. Dans tous ces cas, il n'y aurait pas de perception sans ces actions concomitantes.

Même pour la vision, l'activité est fondamentale. Plusieurs expériences ont été effectuées avec des ordinateurs affichant des images à des sujets, tout en mesurant leurs saccades oculaires. L'une d'entre elles consiste à bouger l'image en même temps que l'oeil saccade, de façon à rendre l'image rétinienne fixe. On constate alors que la vision du sujet devient totalement ineffective! Plus précisément, différents phénomènes se produisent: perte de contraste, fragmentation, champ visuel devenant gris ou noir [7].

Dans une autre expérience, on a demandé aux sujets de discriminer trois symboles chinois inconnus, en faisant en sorte qu'ils pouvaient voir les symboles quand leurs yeux étaient centrés sur eux, mais qu'ils disparaissaient dès que les yeux bougeaient. Les sujets ont trouvé cette tâche extrêmement désagréable et difficile, et ont eu besoin de centaines d'essais pour la réussir [12]. Comme le dit O'Regan : "It is like trying to recognize an object lain on your hand without manipulating it" [13].

Dans notre programme, la part active de l'agent se limite aux aspects de timing: l'agent ne déplace rien et n'agit sur rien, mais il attend les durées précises entre les frappes au clavier. Mais comme nous l'avons montré auparavant, le timing est un aspect essentiel d'une action.

3.3 Les processus perceptifs sont composés de paires action/anticipation

Nous disions: pour sentir une odeur, il faut inspirer. Notre théorie dira plutôt: on inspire, *dans le but de tester* qu'une odeur anticipée est effectivement présente, ou non. On n'inspirerait pas de cette façon, ou à ce moment, si on n'était pas engagé dans cette activité perceptive particulière. On bouge ses doigts *dans le but de vérifier* une sensation anticipée, etc. Même en vision, nous bougeons les yeux pour vérifier des anticipations. Comme le dit O'Regan [13]:

'Red' is the structure of the changes that 'red' causes ... the sensation that a patch of color is red, say, does not primarily derive from the fact that it is stimulating long-wavelength retinal cones. Rather, it generally derives from the *structure of the changes* in sensation that occur when you move your eyes around relative to the patch.

C'est exactement ce que nous essayons de modéliser dans notre programme. Les schémas sensori-moteurs les plus élémentaires ont la structure décrite figure 6.

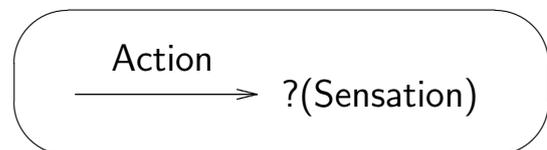


Figure 6: Structure d'un schéma sensori-moteur élémentaire

Cette figure demande à être expliquée. La flèche libellée 'Action' signifie qu'une action est effectuée d'abord, avec son tempo propre. '? Sensation' signifie que, lorsque 'Action' a été réalisée, il faut tester l'occurrence de la 'Sensation' anticipée.

Par exemple, 'Action' = 'inspirer' et 'Sensation' = 'odeur de café'. Ou: 'Action' = 'mouvement oculaire' et 'Sensation' = 'un changement que le rouge cause'. Et notre programme de reconnaissance de rythmes? Il est clair que nos SSMs musicaux sont des éléments chainés de ce type, un pour chaque note dans la partition associée. Clairement, 'Sensation' = 'frappe anticipée'. Mais il semble qu'il n'y ait pas réellement d'action. En fait, 'Action' = 'attendre le temps associé à la note précédente'; il ne s'agit pas d'une action telle qu'un mouvement, mais d'une action d'attente (figure 7).

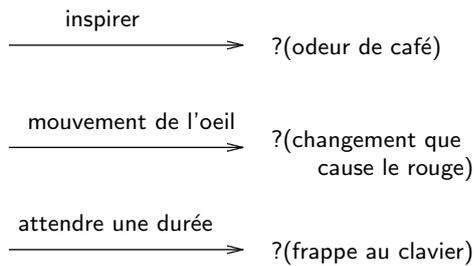


Figure 7: Exemples de schémas sensori-moteurs élémentaires

Appelons E-SSM ces schémas sensori-moteurs élémentaires. Ils sont atomiques, c'est à dire indécomposables. Ils ont les propriétés importantes suivantes:

- un E-SSM est *sensori-moteur*, c'est à dire moteur et sensoriel, les deux aspects étant inséparables
- un E-SSM a une durée intrinsèque d'exécution.
- un E-SSM est normatif : selon l'occurrence de 'Sensation' après 'Action', sa validité peut être mesurée sur une échelle entre 0 et 1.

Dans notre programme, un SSM est une séquence de SSMs atomiques, mais d'autres types de combinaisons sont possibles, et devront être explorés.

3.4 Où se situe la perception dans le cadre cognitif?

Dans la tradition de pensée occidentale, l'intelligence est caractérisée en termes de relations entre idées et environnement. Dans cette perspective, appelée *cognitivism*, la représentation joue un rôle central, en tant qu'unité fondamentale dans la machinerie cognitive. Elle est associée à un modèle de l'esprit de type *perception* → *inférence sémantique* → *action*, où la perception joue le rôle de collecteur d'information. Les aspects physiques du système ne jouent aucun rôle exceptés en tant qu'entrées ou sorties. Ils ne sont jamais intimement associés puisqu'ils sont séparés par la couche *inférence sémantique*.

Cette tradition de pensée a dominé l'intelligence artificielle et les sciences cognitives jusqu'à maintenant, mais ses fondements ont été critiqués ces dernières années par les avocats des approches 'situated cognition' et 'embodied cognition'¹.

Rodney Brooks notamment [5, 4] argumente contre la conception *sense* → *think* → *act* qui n'arrive pas à produire une interaction efficace pour des robots mobiles placés dans un environnement réel. Selon lui, il est trop difficile de construire une représentation du monde suffisamment détaillée pour permettre à un robot d'opérer dans un environnement tel qu'un bureau, et même si c'était possible, le cycle *sense* → *think* → *act* conduirait à des temps de réponse trop longs. A la place, ses collaborateurs et lui ont construit une série de robots se déplaçant autour de leurs bureaux, et capables de d'accomplir des tâches telles que le ramassage de bouteilles de soda vides. Ils ont utilisé une méthodologie appelée *subsumption architecture*, dont les principales caractéristiques sont :

- plusieurs sous-systèmes parallèles actifs répondent à des aspects fonctionnels de la situation, sans se baser sur quelque représentation du monde que ce soit
- pas de contrôle centralisé
- l'environnement est le principal médium de communication entre processus

Donc, selon la conception de Brooks, *l'embodiment* d'un système intelligent est essentiel. A la question : "can there be disembodied mind?", il répond : "what is human about us is very directly related to our physical experiences" [4]. Cette vue est très similaire à celle de Piaget, pour qui toute compétence, même celles qui semblent les plus abstraites, trouve ses racines dans l'intelligence sensori-motrice qui s'est développée durant la période de la petite enfance qui précède le langage. L'inclusion ensembliste, par exemple, a ses racines dans les opérations concrètes analogues sur les objets [16].

D'un point de vue encore plus théorique, le cognitivisme se trouve confronté au *symbol grounding problem*, décrit très clairement par Harnad [10]. Les manipulations de symboles sont purement formelles, et toute tentative de *définir* un symbole ne peut être faite qu'avec d'autres symboles. Cela laisse tous les symboles sans enracinement et donc sans signification. Cette circularité de ne pouvoir exprimer le sens d'un symbole qu'avec d'autres symboles est ce qu'on appelle le *symbol grounding problem* et le cognitivisme ne peut en sortir.

¹Ces approches alternatives existaient depuis longtemps, et notamment en Europe avec Bergson [1], Piaget [14] et Merleau-Ponty [11]

Ainsi, nous croyons que la perception est gouvernée par l'activité de l'agent, sans être un simple collecteur d'information. Toutes les formes d'intelligence sont basées sur l'intelligence sensori-motrice dont la perception est un aspect essentiel.

3.5 Interactivisme

De façon similaire, Mark Bickhard décrit en détails les contradictions de la position cognitive [3]. En particulier, il montre qu'aucune des approches standard de représentation n'est capable de fournir un critère de détection d'erreur *pour l'agent*. Pour cela, le système doit être capable de comparer ce qu'une représentation est supposée représenter, avec ce qui est réellement représenté. Mais déterminer ce qui est représenté est précisément le problème à résoudre, d'où une circularité fatale.

La clé du système de Bickhard consiste à considérer le système dans son interaction dynamique avec l'environnement, et non dans des configurations statiques : en travaillant dans un cadre de processus et non dans un cadre de substance. Dans cette perspective, un système interactiviste a, dans la plupart des situations, plusieurs interactions possibles, et un choix doit être fait pour déterminer vers laquelle il doit s'engager. Pour cela, le système doit posséder des *indications de résultats anticipés* associées à chacune de ces interactions. Ces indications de résultats anticipés doivent être internes à l'agent, c'est à dire lui être directement accessibles. Elles ne doivent pas avoir à être représentées, ou cela conduirait à la contradiction des systèmes symboliques.

Maintenant, l'interactivisme conjecture que ces paires (indications d'interactions potentielles, indications de résultats anticipés) sont les formes fondamentales de représentation et de connaissance, à partir desquelles toutes les autres formes dérivent. Il s'agit d'une connaissance de potentialités, une façon pour l'agent de différencier son environnement, d'accéder à ses propriétés. Ces paires peuvent être pertinentes ou non: les résultats anticipés correspondent ou non à la réalité. Et le point crucial est que ce critère d'erreur est interne à l'agent, qu'il ne dépend pas d'un observateur extérieur pour interpréter ou établir sa valeur de validité.

Il doit être déjà clair que nos schémas sensori-moteurs sont des exemples stricts de représentations interactivistes.

3.6 Les processus perceptifs sont actifs en permanence

Si nous pratiquons un sport intensément durant une journée, des images obsédantes de cette activité vont nous assaillir le soir au moment de dormir. Le même genre d'images se produisent après plusieurs heures de jeu d'échecs, ou après avoir joué à des jeux vidéo.

Sans parler des mélodies musicales, qui peuvent être particulièrement obsédantes.

Toutes ces remarques sont à la fois très banales et très significatives. Elles nous ont inspiré un des principaux points de notre modèle : le fait que les processus SSMS, une fois créés, continuent à être actifs en permanence, en guidant les actions de l'agent. Lorsque nous faisons un tir au football, nous sommes guidés par tous les tirs dans des situations analogues que nous avons faits dans le passé. Et ce sont ces tirs que nous voyons avant de nous endormir, car une longue activité leur a donné une intensité inhabituelle.

Pour supporter ce point de vue, prenons un autre exemple. Lorsque nous écoutons une musique distraitemment, en cuisinant par exemple, nous pouvons être surpris en train de chantonner à certains moments, ou seulement écouter à d'autres. Si le niveau sonore baisse ou si le son devient inaudible par moments, nous restons capables de rester en phase. Nous soutenons que dans tous ces cas, un schéma sensori-moteur est synchronisé sur la musique, et que ses sorties motrices sont plus ou moins inhibées. Mais écouter et chantonner la musique sont deux aspects d'un même processus.

Dans notre programme de reconnaissance, tous les SSMS créés restent actifs en permanence, et guident les actions de l'agent; cet aspect est un des plus caractéristiques de notre modèle.

3.7 Les processus perceptifs sont des créations idéales de l'agent, et non des copies imparfaites de la réalité

Piaget s'est battu toute sa vie contre le problème de la copie dans les systèmes de représentation. (voir par exemple [18]). Son raisonnement est que, pour construire une copie de quelque chose, on doit d'abord connaître ce qui est à copier : encore une autre une circularité fatale.

Pour illustrer cette contradiction, considérons par exemple deux musiques qui sont jouées en face de nous, à des niveaux sonores comparables, et supposons que nous écoutons seulement l'une d'elles, inconnue de nous. Si notre esprit devait faire une copie de cette première musique, afin de la connaître, il devrait éliminer du signal global la seconde, en utilisant en temps réel des opérations d'extraction telles que le filtrage fréquentiel. Mais pour appliquer ces opérations aux bons moments et avec les bons paramètres, notre esprit aura besoin de connaître en détails la première musique. Donc la contradiction est: pour faire une copie d'une musique afin de la connaître, nous devons déjà la connaître.

Si maintenant nous essayons seulement de *discriminer* la première musique, le cercle vicieux précédent devient une spirale vertueuse. Il n'est pas nécessaire de complètement connaître une des

musiques pour être capable de la séparer d'une autre. Si nous sommes capables de distinguer son tempo de base et son rythme (et notre culture musicale générale nous y aidera), ces éléments nous permettront ensuite d'avoir accès à des indices plus fins tels que le rythme détaillé de chaque mesure, la valeur des notes, etc.

Donc, ce qui est construit durant ce mouvement en spirale n'est pas une copie, mais un *processus de discrimination basé sur l'activité et l'anticipation*, et il se construira après plusieurs écoutes et plusieurs raffinements. C'est exactement ce qui se produit dans notre programme. Les SSMs partiellement adaptés sont ces processus de discrimination à des stades intermédiaires, et ils évoluent et se stabilisent finalement pour devenir les processus sensori-moteurs de discrimination optimaux qui justement permettent la reconnaissance.

Les flots sensoriels (visuels, tactiles, phoniques, etc.) ont une bande passante énorme. Quels types de copies pourrions nous y faire? Et à quels moments? Clairement, l'erreur est une fois encore de rester dans une problématique de substance, au lieu d'adopter une problématique de processus.

Un autre point important à propos de ces processus interactiviste est que ce sont des constructions idéales de l'agent. S'ils sont fabriqués à partir d'un flot sensoriel bruité, le bruit n'y sera pas incorporé. Ils peuvent au départ trop simplifier la réalité, mais ils ont une tendance à en incorporer la complexité de plus en plus.

Bien sûr, cette capacité à extraire des informations à partir de sources bruitées a un potentiel d'applications pratiques important.

3.8 Une mémoire de rythmes

Le programme de reconnaissance de rythmes a des capacités évidentes de mémoire à long terme. Il a aussi des formes de mémoire à court terme : lorsqu'un SSM se synchronise sur une musique, il connaît son passé immédiat, et même son futur immédiat ! En même temps, il s'agit aussi d'une mémoire déformante: si l'utilisateur oublie de frapper une touche pendant qu'il joue un air que le programme connaît bien, le programme ne va pas le remarquer.

Dans notre modèle, la mémoire n'est pas une fonction séparée dont le but serait de prendre des photographies (de quoi? quand?). Au contraire, c'est un aspect dynamique de tous les processus cognitifs.

3.9 Le modèle assimilation / accommodation de Jean Piaget

La plus importante intuition du psychologue suisse Jean Piaget est probablement sa notion d'assimilation. Elle lui permet de décrire en détails le comportement de ses trois enfants de leur naissance à 18 mois [15], avec une profondeur d'analyse qui n'a jamais existé

avant et après lui.

En biologie, une entité assimile les nutriments autour d'elle pour maintenir fonctionnelles ses structures internes. Piaget a montré que les mêmes principes sont à l'oeuvre dans le monde cognitif, et qu'il existe en réalité une continuité du monde biologique au monde mental. Les créatures vivantes réagissent aux actions du monde extérieur en fonction de leurs structures mentales, qui en retour s'accommodent à lui dans le but de maximiser cette assimilation. Finalement, elles imposent à l'univers entier un équilibre dépendant de cette organisation.

Nos SSMs sont des schèmes assimilateurs Piagétien : ils incorporent les frappes de l'utilisateur à leur cycle d'activité. Ils s'accommodent à lui de deux façons: une accommodation temporelle fine, qui leur permet d'être synchrones et de continuer à assimiler le flot perceptif, et un processus de diversification / sélection qui permet la création de SSMs de mieux en mieux adaptés.

3.10 Les besoins sont créés par l'activité, et non l'inverse

Ce point a été discuté par Piaget en détails depuis 1934 (date de la première édition française de [15]). La position classique explique que les besoins d'un agent sont le point de départ de son activité mentale. Comment se fait-il que nous répétons sans cesse des comportements utiles tels que chasser, manger, etc.? Parce que, explique ce point de vue, ils satisfont des besoins. Cependant, cette explication pose deux problèmes importants:

1. comment de tels besoins sont-ils capables d'orienter les mouvements et les actions nécessaires à leur satisfaction?
2. comment de nouveaux besoins peuvent-ils être créés?

Il semble difficile de répondre à ces questions dans le cadre précédent, dans lequel les besoins et le comportement ont été séparés.

Essayons de dresser une liste de nos besoins, à nous autres êtres humains. Bien sûr, nous devons satisfaire en permanence certains besoins corporels tels que boire, manger, préserver notre température corporelle, etc. Mais, est-ce que les enfants ont vraiment *besoin* de jouer à des jeux vidéo? Ils semblerait que ce soit le cas : lorsqu'un joueur invétéré ne joue pas pendant quelques temps, il devient nerveux et la plupart de ses actions deviennent orientées vers ce but. Lorsqu'un petit enfant apprend à appuyer sur un interrupteur pour allumer une lampe, il va le faire sur tous les interrupteurs pendant des jours. Il n'y a sûrement pas de besoins innés de jouer aux jeux vidéo ou d'appuyer sur des interrupteurs !

Ainsi, il semblerait que de nouveaux besoins peuvent être créés, et qu'ils puissent ne pas dépendre de processus physiologiques. Dans le cas d'un enfant appuyant sur un bouton, il semble que le besoin ait été créé par l'activité et non l'inverse.

En fait, il n'est pas nécessaire de postuler une séparation entre un besoin et l'activité qui le satisfait. Dès qu'une activité est formée et prend suffisamment d'intensité, un besoin associé de la réaliser s'est créé : c'est le signe d'un déséquilibre temporaire, qui doit être corrigé. Cette activité, composée de schémas sensori-moteurs, oriente les mouvements et les actions de l'agent pour continuer son fonctionnement, et donc elle satisfait le besoin qui lui est associé. La question qui était posée au début du paragraphe est donc résolue, ou plutôt dissoute.

Cette discussion à propos de besoins peut sembler hors de propos dans cet article. Qu'a-t-elle à voir avec la reconnaissance de rythmes? En réalité, elle la gouverne ! Notre programme de reconnaissance a *besoin* de reconnaître des rythmes; toute son activité est orientée dans ce but, et la musique est le principal aliment de cet appétit. Bien sûr, la nature de ces besoins est toujours la même, toujours de reconnaître des classes de rythmes de plus en plus variées. Mais, comme pour tout agent, son activité présente façonne ses besoins et ses actes futures.

4 Conclusion

Notre programme de reconnaissance de rythmes est un programme Java de moins de 400 lignes. Néanmoins, il nous a permis d'implémenter les idées les plus importantes du modèle interactiviste et des schémas d'assimilation à la Piaget.

Un algorithme original, analogue au fonctionnement du système immunitaire des mammifères, a été utilisé pour modéliser le processus de variation / sélection nécessaire à l'accommodation des schémas sensori-moteurs pour les rendre plus finement adaptés à la situation. L'activité permanente en arrière plan de ces schémas a été utilisée pour orienter les actions de l'agent.

Nous avons été capables de mettre en évidence plusieurs aspects importants du modèle interactiviste de l'intelligence : comment le timing et l'anticipation sont des aspects fondamentaux de la cognition, comment ils ont une tendance à s'appliquer chaque fois que c'est possible, comment la représentation peut émerger à partir d'un terrain non représentatif, en utilisant un critère d'erreur accessible à l'agent de manière interne. Nous avons présenté un cadre d'apprentissage d'une nouvelle sorte, qui garde intactes les expériences individuelles de l'agent. Il contraste avec la plupart des algorithmes d'apprentissage basés sur les réseaux de neurones, dans lesquels les expériences passées se mélangent en un ensemble de poids synaptiques.

Des expériences ultérieures sur d'autres modalités sensorielles montreront si notre modèle possède les mêmes propriétés, et continue d'être accessible au calcul.

References

- [1] H. Bergson. *Matière et mémoire*. Paris, 1896.
- [2] M. H. Bickhard. The emergence of representation in autonomous embodied agents. In *Proceedings of the Fall 1996 AAAI symposium on embodied cognition and action*, 1996.
- [3] Mark Bickhard and Loren Terveen. *Foundational issues in artificial intelligence and cognitive science*. Elsevier, North-Holland, New-York, 1995.
- [4] Rodney A. Brooks. Intelligence without reason. In *International Joint Conference on Artificial Intelligence*, pages 569–595, 1991.
- [5] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–160, 1991.
- [6] E.J. Coyle and Shmulevich I. A system for machine recognition of music pattern. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1998.
- [7] R. W. Ditchburn. *Eye-movements and visual perception*. Clarendon Press, Oxford, 1973.
- [8] S. Forrest and S. Hofmeyer. Immunology as information processing. In L. Segel I. Cohen, editor, *Design Principles for the Immune System and other Distributed Autonomous Systems*. Oxford University Press, 2000.
- [9] J. Gibson. The problem of temporal order in stimulation and perception. *Journal of Psychology*, 62:141–149, 1966.
- [10] Stevan Harnad. The symbol grounding problem. *Physics D*, 42:335–346, 1990.
- [11] M. Merlau-Ponty. *Phénoménologie de la perception*. Paris, 1962.
- [12] T. A. Nazir and J. K. O'Regan. Some results on translation invariance in the human visual system. *Spatial Vision*, 5:2:81–100, 1990.
- [13] J. Kevin O'Regan and Alva Noë. A sensori-motor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24:5, 2001.
- [14] J. Piaget. *La naissance de l'intelligence chez l'enfant*. Delauchaux et Niestlé, Neuchâtel, 1934.
- [15] J. Piaget. *La construction du réel chez l'enfant*. Delauchaux et Niestlé, Neuchâtel, 1936.

- [16] J. Piaget. *Biologie et connaissance*. Galimard, Paris, 1967.
- [17] J. Piaget. *Genetic Epistemology*. Columbia, New-York, 1970.
- [18] L. A. Segel. The immune system as a prototype of autonomous decentralized systems. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, 1997.
- [19] B. Shannon. *The representational and the presentational*. Harvester Wheatsheaf, 1993.
- [20] I. Shmulevich and D. Povel. Measures of temporal pattern complexity. *Journal of New Music Research*, 29(1), 2000.