Lire dans le regard des robots

Gérard Bailly et **Frédéric Elisei,** Gipsa-lab, CNRS, université Grenoble Alpes et Grenoble INP

Les robots sont de plus en plus performants. Mais ils sont encore loin de maîtriser toutes les subtilités d'une interaction humaine. Au Gipsa-lab, à Grenoble, des roboticiens leur enseignent comment donner à voir leurs intentions en leur apprenant à ajuster leurs comportements – gestes, parole, regard – en fonction du contexte et de leur interlocuteur.

ROBOTICIENS
Gérard Bailly (1) est directeur de recherche CNRS au Gipsa-lab, où il dirige l'équipe Cognitive robotics, interactive systems & speech processing (Crissp). Frédéric Elisei (2) est ingénieur de recherche CNRS au sein de cette équipe.

Dans son guide interactif *Mind Reading*, publié en 2004, le psychologue britannique Simon Baron-Cohen distingue plus de 400 émotions pouvant être véhiculées par la voix et les expressions faciales : irritation, déception, indignation, soulagement, euphorie, etc. Si tout le corps, par les postures et les gestes, participe à exprimer nos états mentaux, le visage est particulièrement informatif. Les yeux et la bouche font partie des zones les plus « lues » par nos interlocuteurs. Des études montrent même que la morphologie de l'œil humain a évolué de manière à maximiser la précision du décodage de la direction du regard par autrui (lire l'encadré p. 68). Cet œil « coopératif » facilite ainsi le développement d'une théorie de l'esprit (*) évoluée chez les êtres sociaux que nous sommes devenus.

Sentiment d'étrangeté

Quid des machines ? Nombre de robots humanoïdes n'ont pas de visage visible (Asimo, Jibo), d'autres en possèdent un, mais il est fixe (Nao, Pepper) ou a été remplacé par un écran plat (Baxter). En revanche, certains d'entre eux, comme Sophia, mis au point par la société Hanson Robotics, ou les geminoïdes conçus par le roboticien japonais Hiroshi Ishiguro, sont construits avec un grand réalisme physique. Leur visage est fait de peau synthétique et leurs yeux sont dotés de cristallins oculaires imitant les nôtres.

Paradoxalement, cette ressemblance peut être une source de déception pour l'interlocuteur, qui s'attend à une interaction d'autant plus humaine que le robot est réaliste. Or il est techniquement difficile de reproduire sur une machine les 38 muscles de la face et leurs entrelacements complexes. Résultat : quand le squelette sous-cutané des robots humanoïdes du type de Sophia n'a pas suffisamment de degrés de liberté pour reproduire des mouvements humains, en particulier au niveau du visage, ils nous procurent un sentiment d'étrangeté (lire l'encadré p. 69). Quant aux autres robots humanoïdes, dont la plastique est moins réaliste, la plupart n'ont pas d'œil coopératif fonctionnel : leurs caméras

Contexte **En psychologie, la théorie de l'esprit définit notre capacité à lire les états mentaux d'autrui. Cela passe par des signaux non verbaux, comme les expressions faciales. En enseignant aux robots sociaux la maîtrise de ces signaux, nous pourrions interagir plus naturellement avec eux. Cette recherche pionnière vise des applications dans le champ médical et en robotique collaborative.**

DR

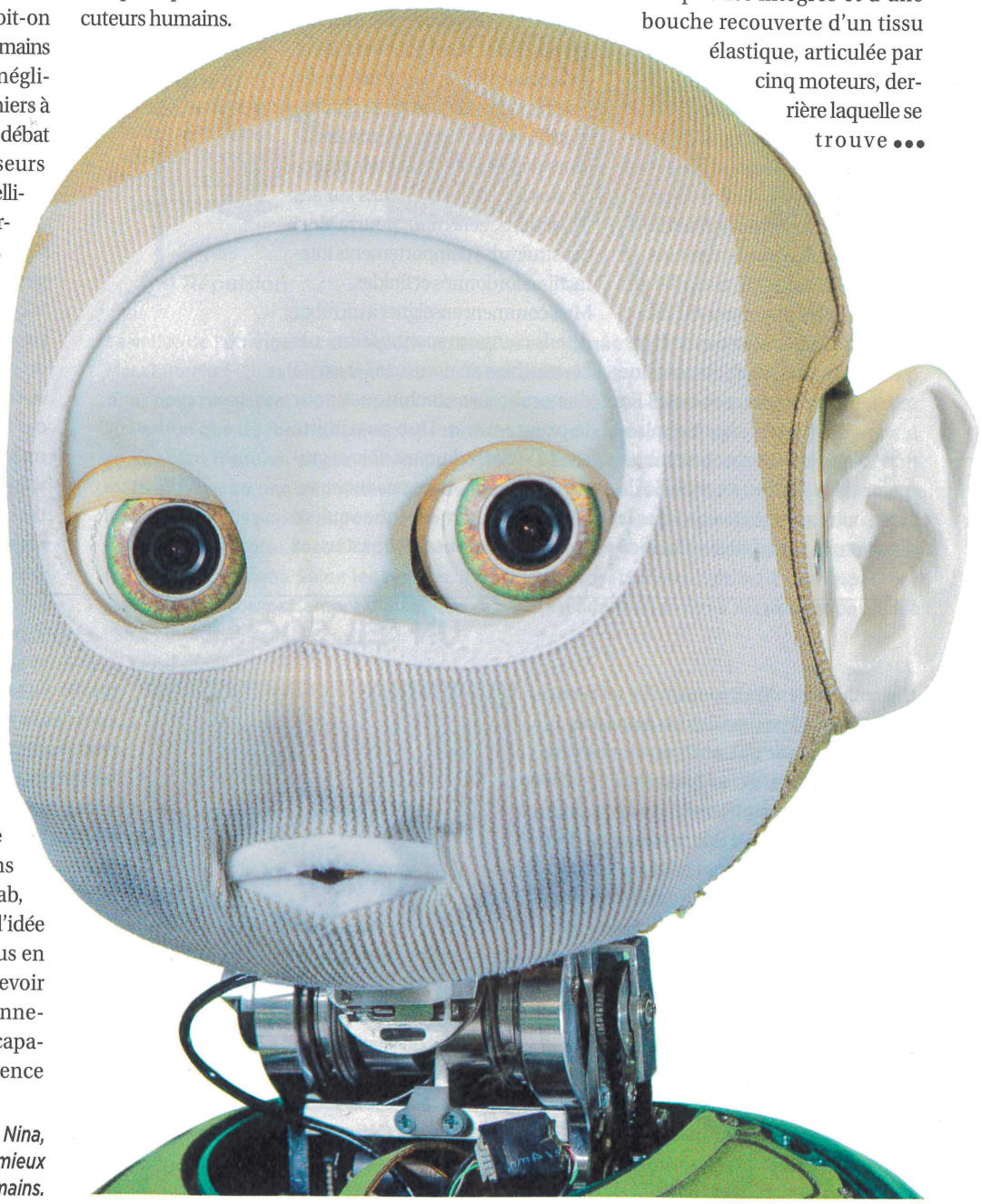
sont souvent fixes, incrustées au milieu du front ou sur le torse. Leurs concepteurs les munissent par ailleurs de capteurs – caméras 2D et 3D, télémètres et radars, microphones… – qui mesurent et analysent nos moindres faits et gestes, mais ils oublient souvent qu'ils devraient, symétriquement, « donner à voir » les intentions de leurs créatures.

C'est là que le bât blesse ! Doit-on maintenir l'asymétrie entre humains et robots en limitant ou en négligeant la capacité de ces derniers à exprimer leurs intentions ? Le débat est ouvert entre les défenseurs d'une humanité ou d'une intelligence sociale supérieure réservée aux seuls humains, et les promoteurs d'un transhumanisme débridé par les performances surhumaines de l'intelligence artificielle dans des habiletés motrices – locomotion, saisie d'objets au vol, etc. – mais aussi cognitives – calcul, mémoire, etc. Notre positionnement, plus nuancé, est de donner aux robots la capacité de remplir des fonctions identiques à celles des hommes sans nécessairement en faire des copies conformes. Ces fonctions peuvent concerner la locomotion et la manipulation d'objets ou, dans notre cas, la communication verbale et l'interaction sociale. Dans notre laboratoire, le Gipsa-lab, à Grenoble, nous défendons l'idée que, si les robots sont de plus en plus performants pour percevoir et comprendre leur environnement, il faut leur donner la capacité de partager leur expérience avec nous, et ceci sans avoir à réapprendre de nouveaux codes de communication.

La recherche dans ce domaine n'en est qu'à ses balbutiements. Même si les robots agissent sur ce qui les entoure, peu d'études se sont attachées à vérifier que leurs actions – notamment communicatives – sont correctement perçues et acceptées par leurs interlocuteurs humains.

(*) **La théorie de l'esprit** est la capacité à se représenter les états mentaux d'autrui.

C'est ce que nous avons voulu tester sur notre robot humanoïde, Nina, mis au point en partenariat avec l'Institut italien de technologie, à Gênes. Nina est une version augmentée du robot iCub, conçu en 2006 pour étudier la cognition robotique. Elle est ainsi dotée d'un nouveau mécanisme d'articulation des paupières, d'oreilles avec microphones intégrés et d'une bouche recouverte d'un tissu élastique, articulée par cinq moteurs, derrière laquelle se trouve •••



▶ *Grâce à ses yeux articulés, Nina, le robot du Gipsa-lab, peut mieux se faire comprendre des humains.*

CYRIL FRESILLON / GIPSA-LAB / CNRS PHOTOTHÈQUE

DR

••• un haut-parleur qui lui donne de la voix. Les mouvements de sa bouche sont commandés par un synthétiseur de parole développé par notre laboratoire. Ce système, entraîné au préalable grâce à des vidéos de discussions humaines, est capable de calculer et d'exécuter des mouvements réalistes de la bouche et du visage tout en diffusant les paroles correspondantes de façon synchrone (1). Grâce à des tests réalisés dans un environnement bruyant, en l'occurrence le fond sonore d'une soirée cocktail, nous avons montré que les mouvements de la mâchoire et des lèvres calculés par notre système amélioraient bien la perception audiovisuelle d'interlocuteurs humains (2). Nous avons ensuite travaillé sur les yeux et les paupières du robot, en nous posant cette question : quels paramètres influencent une meilleure interaction sociale ? En testant différentes capsules plastiques autour des caméras embarquées dans les yeux articulés de Nina, nous avons démontré que la direction du regard était estimée de manière plus précise par ses interlocuteurs humains lorsque les tailles relatives de la sclère (*) blanche et des iris colorés étaient judicieusement choisies et que l'abaissement des paupières supérieures accompagnait celui du regard (3). Quand Nina observe un objet qui l'intéresse, son interlocuteur peut donc le savoir. Or cette capacité à diriger son regard de façon à ce que l'interlocuteur en saisisse la cible est l'une des bases de l'attention partagée et de la construction d'une théorie de l'esprit. Grâce à ce partage d'informations implicites, les interlocuteurs de Nina vont pouvoir élaborer des hypothèses fiables sur ses (ré)actions, et le robot pourra ainsi construire des comportements interactifs coordonnés et fluides.

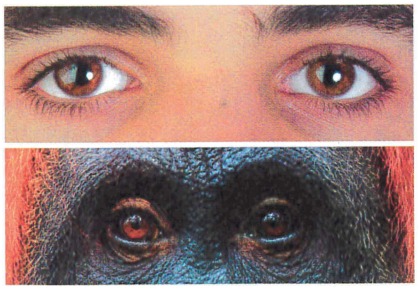
Mais comment enseigner à un robot de tels comportements pertinents et sensibles au contexte ? Il existe de nombreuses techniques pour le programmer. Une possibilité est de le doter d'un modèle cognitif lui permettant de raisonner sur son environnement (que veulent les interlocuteurs ? Qui est assez proche pour saisir tel objet ? etc.) et de réagir aux ordres, aux questions, aux affirmations, aux doutes d'autrui, de manière à planifier ses propres actions. Cette approche commence à être complétée, voire supplantée, par des techniques d'apprentissage statistique et d'intelligence artificielle permettant de faire correspondre le flux des signaux perçus par le robot (parole, gestes…) avec des actions à exécuter (observer telle zone, désigner tel objet, prononcer tel mot…). En bref, ces techniques permettent de capturer les régularités des comportements interactifs.

Parmi les diverses techniques d'apprentissage statistique qui peuvent s'appliquer, la première est l'apprentissage dit « développemental » : le robot apprend tout seul par essais-erreurs, ce qui demande un nombre prohibitif d'essais. En outre, l'algorithme utilisé dans ce cas nécessite de définir explicitement ce qu'est un bon (ou un mauvais) comportement social interactif. Or il n'existe pas de définition universelle pour en juger. Une

(*) **La sclère** est la membrane blanche et opaque qui forme le blanc de l'œil. Selon les espèces, elle est plus ou moins pigmentée.

UN ŒIL SOCIAL

L'hypothèse de « l'œil coopératif » a été proposée en 2001 par une équipe japonaise (1), puis reprise par des anthropologues de l'Institut Max-Planck, en Allemagne (2). Elle avance que la morphologie et l'apparence de l'œil humain – sclère blanche en fort contraste avec l'iris et la peau du visage – ont évolué de manière à faciliter la lecture de la direction de notre regard par autrui et ainsi favoriser les activités coopératives. Pour le montrer, les chercheurs ont comparé la forme et l'apparence de l'œil (le rapport hauteur/largeur et la surface exposée de la sclère) de 874 animaux adultes de 88 espèces, le contraste colorimétrique entre la sclère, l'iris et la peau de 92 espèces, et les mouvements oculaires de 26 espèces.



▲ *Parmi 92 espèces étudiées, les humains sont les seuls à posséder une sclère blanche.*

Résultat : l'œil humain est celui dont la sclère est la plus exposée et dont l'élongation horizontale est la plus importante parmi tous les primates. Cette morphologie favorise des mouvements oculaires plutôt qu'une rotation de la tête : 61 % des explorations visuelles se font juste en bougeant les yeux chez les humains, contre 20 % à 35 % chez les chimpanzés. Par ailleurs, 85 des 92 espèces observées présentent une sclère de couleur marron, et seuls les humains en ont une blanche. Mieux : nous sommes les seuls dont la sclère est plus claire que la peau et l'iris. Selon les auteurs, la pigmentation de l'œil présenterait un avantage sélectif : camoufler le regard, et donc les intentions. Chez l'homme, il est probable que ce trait ait disparu en faveur d'un regard facilitant la communication des intentions.

(1) H. Kobayashi et S. Kohshima, *J. Hum. Evol.*, *40*, 419, 2001.

(2) M. Tomasello *et al.*, *J. Hum. Evol.*, *52*, 314, 2007.

ADAM FRAISE/SHUTTERSTOCK -

autre option est l'apprentissage par observation ou par imitation. Pour un robot, cela consiste à examiner un tuteur humain pendant qu'il exécute une tâche avant de la reproduire. Problème : le robot n'est pas aussi agile qu'un humain – les mouvements de notre visage sont bien plus riches et divers que ce que peut produire un robot. Il doit donc transposer les réactions humaines en les adaptant à ses capacités sensorimotrices (et cognitives) limitées. Ensuite, il faut supposer que ses futurs interlocuteurs réagiront de la même manière face à lui que face à un tuteur humain. Rien n'est moins sûr : même si nous ne pouvons pas nous empêcher de projeter des facultés cognitives humaines sur un dispositif doté

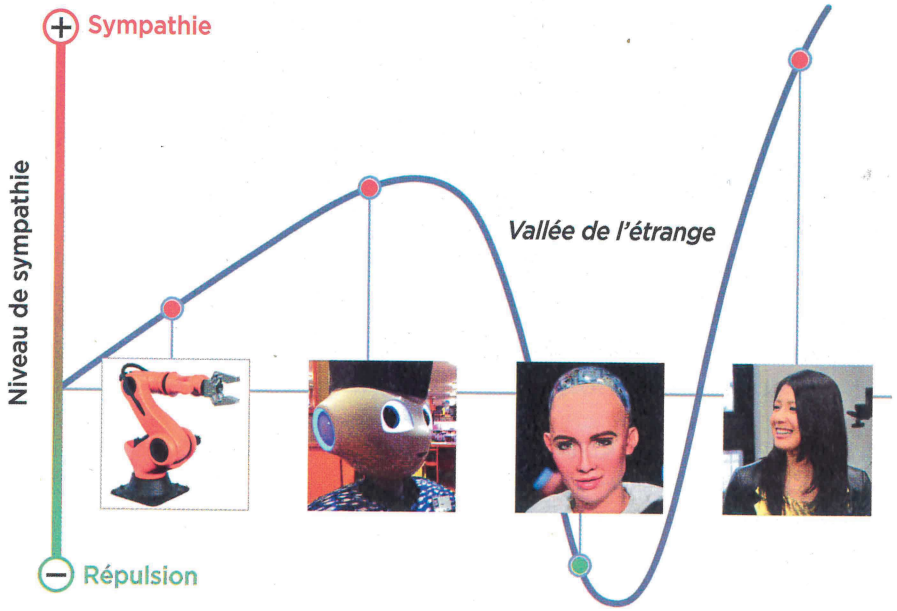
Notre ambition est d'apprendre à Nina à conduire un test de dépistage de la maladie d'Alzheimer

d'initiative, le robot reste un objet technologique. Or notre théorie de l'esprit apprend à séparer les agents (qui agissent sur le monde et ont des intentions) des objets (qui subissent passivement les actions des agents et des forces de la nature) ; nous attribuons à chacune de ces catégories des systèmes de valeurs distincts qui, dans le cas des robots, entrent en conflit. Il est donc difficile de transposer les comportements observés entre humains à ceux qui sont attendus d'un robot social.

Une dernière approche, l'apprentissage par démonstration, permet de prendre en compte les contraintes sensorimotrices de la machine en laissant la possibilité au tuteur d'agir directement sur les actionneurs (les moteurs et les articulations) du robot à la manière d'un marionnettiste. Pour cela, le robot est réglé sur un mode passif ou « docile », dans lequel il suit les gestes du pilote humain et se limite à compenser l'impact du poids de son propre corps sur ses mouvements.

Pour inculquer à Nina des comportements socio-communicatifs, nous avons choisi cette dernière option. Nous avons couplé le robot à une plateforme de « téléopération immersive », où la démonstration est faite de l'« intérieur ». Concrètement, grâce à un casque de réalité virtuelle équipé de dispositifs de capture de mouvements, dont un oculomètre binoculaire, le tuteur, devenu « pilote », agit et perçoit via le corps robotique de Nina et ses capteurs. Nina suit alors passivement les comportements du pilote. Elle stocke dans sa mémoire l'ensemble des signaux sensori-moteurs vécus lors d'expériences d'interactions entre le pilote et des interlocuteurs en situation face au robot. Une fois que cette mémoire comportementale dispose d'un ensemble suffisant •••

LA VALLÉE DE L'ÉTRANGE



La vallée de l'étrange est une théorie scientifique imaginée en 1970 par le roboticien japonais Masahiro Mori (1), selon laquelle plus un robot androïde nous ressemble, plus ses imperfections nous paraissent monstrueuses. Ce n'est qu'au-delà d'un certain degré de réalisme dans l'imitation que les robots humanoïdes seraient les mieux acceptés. Cette vallée fait référence au concept freudien de *unheimlich*, un événement qui a lieu dans une situation familière mais qui suscite une angoisse, voire de l'épouvante. Plus récemment, des chercheurs de l'université d'Osaka, au Japon (2) ont proposé une cartographie plus complexe, prenant en compte nos attentes liées à l'apparence statique du robot et nos impressions résultant de sa mise en mouvement. Selon les auteurs, il existerait une « colline » optimale où l'apparence statique et le comportement dynamique seraient en adéquation et ne mettraient pas mal à l'aise l'interlocuteur.

(1) M. Mori, *Energy*, 7, 33, 1970.

(2) T. Minato *et al.*, in IEA-AIE 2004, *Innovations in Applied Artificial Intelligence*, Springer, p. 424, 2004.

▲ *Nina conduit des entretiens neuropsychologiques avec des humains.*

••• d'exemples d'interactions, des techniques d'analyse de données et des modèles statistiques sont utilisés afin de doter Nina de comportements interactifs autonomes.

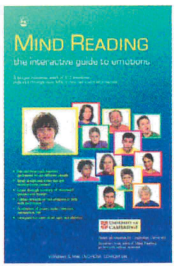
Courtes interactions

Dans le cadre du projet Sombrero, financé par l'Agence nationale de la recherche, notre ambition est d'apprendre à Nina comment conduire des entretiens neuropsychologiques de manière autonome (4). Ces derniers consistent à évaluer la mémoire épisodique (*) de patients pour lesquels il existe une suspicion de maladie d'Alzheimer, ou d'autres types de démence, en leur faisant passer un test standard de seize items. En temps normal, ces entretiens individuels sont menés par des médecins. Ce type d'échange court – un examen prend environ vingt minutes – est bien adapté à ce que l'on peut confier à un robot social. C'est en effet une tâche répétitive, avec un protocole standardisé, donnant lieu à une interaction finalisée où les rôles sont bien déterminés. Il ne s'agit pas de se substituer au praticien pour faire un diagnostic, mais de filtrer les patients qui passent le test robotisé avec succès pour dépister ceux qui doivent consulter un spécialiste. Il ne faut cependant pas sous-estimer une difficulté : s'adapter aux milliers de profils psychologiques des candidats au dépistage, en maintenant une attention constante, une bienveillance courtoise et une empathie infaillible. Tâche laborieuse mais ô combien utile quand il s'agit d'offrir un dispositif fiable et discret – une consultation robotique pouvant être perçue comme moins « impliquante » que celle d'un médecin. Un tel dépistage précoce bénéficierait au nombre croissant de personnes malades, dont plus de la moitié n'est actuellement pas diagnostiquée.

Contrairement aux robots compagnons pour lesquels la difficulté (éthique et technique) consiste à construire une relation à long terme avec un seul humain, notre robot social doit avoir des interactions courtes avec une multitude de patients. Il existe de nombreux scénarios d'interactions courtes : du robot intervieweur au robot collaboratif industriel, en passant par l'animateur de jeux ou d'accueil en magasin. L'enjeu est donc de faire en sorte que le robot soit capable de paramétrer rapidement des modèles pré-appris pour les adapter aux réactions motrices, perceptives et cognitives des humains auxquels il s'adresse, et desquels il ne connaît rien ou presque. À l'heure où les avancées rapides de l'intelligence artificielle et de la robotique suscitent enthousiasme, crainte et fantasme, le roboticien peut apporter des outils technologiques complémentaires au débat. En particulier, il peut inculquer au robot la connaissance des limites de ses propres compétences. À l'instar de l'agent conversationnel Tay mis en ligne sur Twitter par Microsoft en 2016, tout système interactif est par définition influencé par ce qu'il perçoit… ce qui peut le conduire à déraper hors de son domaine de compétences, s'il est sollicité par un usager malintentionné ou s'il surestime sa capacité à recadrer l'interaction. En 24 heures, Tay était ainsi devenu raciste et misogyne, obligeant Microsoft à le déconnecter.

Au robot donc de savoir identifier s'il est capable – ou non – de traiter une entrée perceptive et d'estimer s'il est toujours dans son domaine de compétence sociocommunicative. D'où l'importance d'équiper notre robot social d'un « bouton rouge », logiciel dont nous, roboticiens et autres spécialistes de l'interaction sociale ou cognitive, avons la responsabilité. Ce dernier doit permettre au robot d'estimer automatiquement si le modèle d'interaction dont il est doté est capable de gérer un échange avec un humain et de savoir répondre « je ne sais pas » ou de revenir à un état de repos lorsque ce modèle est trop éloigné de la tâche pour laquelle il a été entraîné. ■

(*) **La mémoire épisodique** concerne les souvenirs autobiographiques.



POUR EN SAVOIR PLUS

- Simon Baron-Cohen, *Mind Reading*, Jessica Kingsley Publishers, 2004.
- **www.gipsa-lab.fr/projet/SOMBRERO/videos.html** La présentation du projet Sombrero, coordonné par Gérard Bailly.

English version

Cet article est disponible en anglais sur researchinfrance.com

(1) G. Bailly *et al.*, *Eurasip JASMP*, doi:10.1155/2009/769494, 2009.

(2) A. Parmiggiani *et al.*, *Int. J. Human. Robot.*, *3*, 1550026, 2015.

(3) F. Foerster *et al.*, IEEE-RAS, 15th International Conference on Humanoid Robots, 2015.

(4) D.-C. Nguyen *et al.*, *Pattern Recognit. Lett.*, *100*, 29, 2017 ; D.-C. Nguyen *et al.*, *IEEE CogInfoCom*, 337, 2016.

AFP PHOTO / JEAN-PIERRE CLATOT

DR