Des clones virtuels parlants pour mieux comprendre le langage

Des chercheurs grenoblois ont construit un modèle numérique articulé pour analyser les mécanismes de la parole. Ce visage artificiel permet d'étudier les caractères physiologiques associés à la formation et à la structure des langues

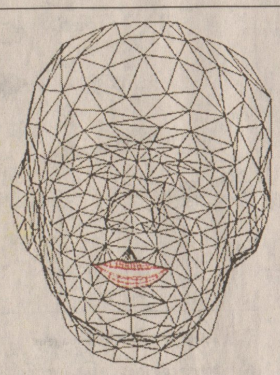
GRENOBLE
de notre envoyé spécial

Pierre Badin a fait don de son visage à la science. De son visage et de sa voix. Ce chercheur de l'Institut de la communication parlée (ICP, CNRS-Institut national polytechnique de Grenoble-université Stendhal) s'est fait filmer sous toutes les facettes, a été radiographié sous tous les angles, a porté des appareillages de toutes sortes. Il a récité des chapelets de *« a-ba-i-bi-ou-bou »*, *« a-da-i-di-ou-dou »* ou encore *« a-ga-i-gi-ou-gou »*. Des années d'un patient labeur pour cet étonnant résultat : un clone parlant. Un double presque parfait. Non pas une réplique réelle, bien sûr. Mais un alter ego virtuel, un sosie numérique doué de la parole. Comme un reflet faisant écho, de l'autre côté de l'écran de l'ordinateur, à l'original. Mêmes traits, même voix, mêmes mimiques.

Cette création, qui pourrait sembler être l'expression d'une crise de narcissisme aiguë, s'inscrit en fait dans les recherches que mène le laboratoire grenoblois, fédérant linguistes, phonéticiens, physiciens, acousticiens, cogniticiens et spécialistes du traitement du signal, sur la production et la perception de la parole. *« On assimile volontiers la parole à un simple signal acoustique. Mais elle est en fait un ensemble de gestes rendus audibles et visibles »*, observe Gérard Bailly, directeur de recherches à l'ICP. Ce qui permet par exemple aux malentendants de lire sur les lèvres. Or cette union intime du sonore et du visuel fait appel à une complexe machinerie corporelle. Celle-ci met en action, outre les poumons et les cordes vocales, pas moins de cinq « articulateurs » – mâchoire, langue, lèvres, voile du palais, larynx – mus par plusieurs dizaines de muscles. Un mécanisme trop sophistiqué pour être reproduit en laboratoire et que l'équipe de l'ICP s'est employée à simplifier.

Son « cobaye », Pierre Badin, s'est soumis à une batterie de mesures faisant appel à différentes techniques, aucune ne permettant à elle seule d'analyser, d'un double point de vue spatial et temporel, la gestuelle de la parole. Le chercheur a d'abord collé sur son visage quelque 300 billes minuscules, comme autant de repères dont les déplacements ont été filmés tandis qu'il prononçait un échantillon d'une cinquantaine de sons élémentaires et dont les coordonnées spatiales ont été enregistrées une à une. Le procédé, dit de capture de mouvement, est courant – c'est celui qui a donné vie à la séduisante créature virtuelle Eve Laura Solal (*Le Monde* du 28 décembre 2000) –, mais il a été poussé ici à un degré de précision extrême, avec un total de 60 000 positions répertoriées. Notre homme est ensuite passé par la cinéradiographie, qui a permis d'établir plus de 2 000 images de ses articulateurs en action, puis par l'imagerie par résonance magnétique (IRM), afin d'obtenir des vues tridimensionnelles des mêmes organes. Enfin, il s'est livré à des séances d'articulographie pour enregistrer les mouvements de sa langue grâce à des capteurs magnétiques miniaturisés.

> « D'un point de vue mécanique, rien ne s'oppose à ce que le singe produise de la parole »



Les mouvements du visage du locuteur sont repérés à l'aide des déplacements de petites billes, dont les coordonnées spatiales permettent de construire un modèle numérique maillé.

Le modèle est recouvert d'un masque, par translation d'une image du locuteur. Les mouvements des machoires, de la langue, du voile du palais et du larynx peuvent être visualisés en rendant le masque transparent.

ICP

Munis de ces multiples données croisées, les chercheurs ont réussi à identifier une dizaine de mouvements fondamentaux – ou degrés de liberté – associés à la locution : deux pour la mâchoire (ouverture-fermeture, avancée-rétraction), trois pour les lèvres (étirement-arrondissement, ouverture-fermeture, élévation-abaissement), quatre pour la langue (avant-arrière, tendue-comprimée, montée-descente, avancée-rétraction de la pointe), un pour le voile du palais (ouverture-fermeture), un dernier pour le larynx (élévation-abaissement). Ils ont pu alors construire un modèle de tête parlante extrêmement réaliste, véritable jumeau numérique du locuteur. Enfin, pour parfaire l'impression de véracité, ils ont donné au modèle maillé une texture, en le revêtant d'un masque transposé, point par point, d'une image du visage réel de Pierre Badin. Ce dernier n'est du reste plus seul dans la galerie de clones de l'ICP. Trois autres modèles, l'un de femme, les deux autres de locuteurs de langue allemande et arabe, ont été réalisés.

Pourquoi tant d'efforts ? *« Les clones articulés constituent un outil privilégié de recherche sur la structure des langues, sur l'émergence du langage chez l'homme ou sur son développement chez l'enfant »*, explique Gérard Bailly. Ces auxiliaires vont donc aider le laboratoire grenoblois à mettre en évidence les liens existant entre la structure des langues – leurs systèmes vocalique et consonantique, leur organisation syllabique – et les capacités physiologiques de production et de perception du langage. Ils l'aideront également à repérer des *« constantes »* résistant à la diversité des 5 000 à 6 000 idiomes répertoriés dans le monde.

L'ICP a déjà montré que *« les langues n'ont pas choisi leurs voyelles arbitrairement »*. Celles qui ne possèdent que trois unités vocaliques, par exemple, ont presque toutes sélectionné le « i », le « a » et le « ou ». Ces trois sons, qui exploitent au mieux les capacités des appareils phonatoire et auditif humains, sont, de ce fait, les plus faciles à distinguer. Pour la plupart, les langues à cinq voyelles ajoutent à cette triade le « è » et le « o », le « u » n'apparaissant que dans les langues à sept voyelles.

N'y a-t-il pas, à considérer la parole comme *« un produit biologique de l'espèce humaine »*, le risque d'occulter sa dimension culturelle ? Les chercheurs estiment au contraire que l'étude de la manière dont l'homme exploite ses facultés langagières *« permet de mieux comprendre les capacités cognitives mises en jeu dans cette activité »*. Justifier par l'anatomie la primauté de quelques voyelles ne dispense pas d'expliquer pourquoi certaines langues, comme l'arabe, ont recours à trois voyelles seulement, quand des dialectes africains en comptent vingt-quatre !

L'institut apporte même une contribution au débat, très actuel, sur la différenciation entre l'homme et l'animal. L'idée très largement admise selon laquelle les singes n'ont pas accès au langage articulé en raison de la position de leur larynx, trop haut placé, est battue en brèche par les simulations. *« D'un point de vue mécanique, rien ne s'oppose à ce que le singe produise de la parole »*, indique Louis-Jean Boë, qui étudie cette question. L'avantage de l'homme sur le singe serait donc lié à ses seules aptitudes cognitives. Paradoxalement, c'est le clone virtuel qui révèle ainsi la réalité de notre humanité.

Pierre Le Hir

Des applications en orthophonie, chirurgie ou télécoms

LES CHERCHEURS de l'Institut de la communication parlée (ICP) de Grenoble ont des projets plein leurs cartons pour leurs clones parlants. Ces derniers pourraient servir d'assistants en orthophonie en aidant les patients à visualiser les mouvements articulatoires du langage. La position de la langue, en particulier, pourrait être mise en évidence par transparence du masque numérique. Ces auxiliaires seraient aussi mis à profit pour l'apprentissage des langues étrangères, en montrant comment former des sons inconnus dans la langue maternelle du locuteur, comme le « u » sur lequel butent, par exemple, les Japonais. Les scientifiques étudient également, avec des spécialistes de la chirurgie maxillo-faciale, l'utilisation de ces modèles à des fins de chirurgie prédictive. Avant de pratiquer une intervention (ablation de la langue, greffe, correction à but esthétique), le praticien ferait voir et entendre à son patient les conséquences de l'opération sur ses capacités d'élocution afin de déterminer l'acte chirurgical le mieux approprié.

UN PROCÉDÉ ADAPTABLE À LA VIDÉOCONFÉRENCE

A un horizon moins lointain, les clones articulés pourraient trouver un emploi dans les télécommunications. L'ICP est engagé, avec France Télécom, Philips et STMicroelectronics, dans un projet de visiophonie où les paramètres des mouvements des lèvres du locuteur, filmés par une microcaméra, sont transmis à un modèle de visage affiché sur l'écran du téléphone portable du récepteur. Le système divise par cent ou mille le débit d'informations nécessaire par rapport à la transmission d'images réelles. Le même procédé est adaptable à la vidéoconférence, en substituant aux images des interlocuteurs réels leurs clones virtuels. Autre projet, conduit avec Arte, Thalès, la société Attitude Studio (créatrice d'Eve Solal) et l'université de Compiègne : le remplacement des textes et des sous-titrages des émissions télévisées par des visages animés, sur les lèvres desquels liraient les malentendants.

Cette approche diffère de celle du 3D Facelab de France Telecom, à Rennes, également lié avec l'ICP. Alors que le moteur d'animation des Grenoblois s'appuie sur le rendu de la texture du visage, les avatars produits à Rennes cherchent à multiplier les triangles qui forment chaque facette du visage – entre 3 000 et 15 000, animés en temps réel et en trois dimensions – tout en restant compatible avec une animation sur la Toile. *« Notre moteur permet d'adapter le nombre de triangles à la puissance d'un ordinateur grand public »*, explique Gaspard Breton, doctorant au Facelab.

Le système permet de créer rapidement des avatars à partir de deux photos (face et profil). Il suffit de sélectionner quelques points d'attache sur la machoire et le cou pour que les vingt-cinq muscles du modèle animent le visage. *« Il agissent comme des aimants attirant les points du maillage dans leur champ d'influence »*, explique le chercheur. De quoi prendre en compte la « paresse » musculaire ou la coarticulation et éviter l'agitation désordonnée des lèvres de certains clones. Chaque expression sollicite en moyenne vingt-cinq paramètres et trente-quatre phonèmes différents peuvent être automatiquement produits à partir d'un texte. Mais pour l'heure, peu d'applications ont vu le jour.

P. L. H. et H. M.

Les créatures artificielles affirment leur personnalité

MONACO
de notre envoyé spécial

Katherine présente des informations médicales à partir de dépêches d'agences de presse. Lors du festival Imagina.02 qui s'est tenu à Monaco du 12 au 14 février, ce personnage virtuel s'est exprimé avec l'aisance d'une professionnelle de la télévision. De plus, à la différence de ses collègues humains, il est possible de lui poser des questions afin, par exemple, d'obtenir des précisions sur l'une des informations qu'elle vient de donner. Barbara Hayes-Roth, professeur au département informatique de l'université américaine Stanford, en Californie, et fondatrice de l'entreprise Extempo, va même jusqu'à l'agresser : *« Tu n'es qu'une création virtuelle stupide... »* A ces mots, Katherine rétorque qu'elle est loin d'être stupide. Pour preuve, son aptitude à exprimer une opinion personnelle sur les sujets qu'elle traite. De fait, lorsqu'on lui demande son avis sur les bienfaits du chocolat, elle l'exprime sans hésiter et sans masquer ses doutes au sujet du sérieux de certaines annonces médicales... A la fin de la conversation, lorsque son interlocuteur prend congé, Katherine démontre que, si elle n'est pas rancunière, elle n'oublie pas. *« Je vous souhaite une bonne journée, malgré vos propos désagréables... »*

DISSÉQUER LES MIMIQUES

Barbara Hayes-Roth a analysé les caractéristiques des créatures artificielles capables de favoriser leurs relations avec le monde réel. Pour éviter les *« pertes de crédibilité »* souvent observées, elle suggère de bâtir des personnages virtuels qui ne se limitent pas à répondre à des questions ou à donner des informations. La recherche d'un enrichissement des relations sociales la conduit à doter ses créations d'une personnalité, d'une vie passée et d'opinions qui permettent d'étendre l'échange avec les humains à des considérations plus intimes.

Au fameux Media Lab du MIT, Justine Cassel suit une démarche parallèle. En visionnant des centaines d'heures de vidéo, son équipe a disséqué toutes les mimiques, les gestes et les intonations utilisés par les personnes réelles lorsqu'elles discutent entre elles. Justine Cassel souligne l'importance des bavardages et des confidences dans ce contexte. L'étude du comportement d'un agent immobilier démontre ainsi qu'il vend plus facilement une maison s'il s'enquiert de la situation familiale de ses clients avant d'aborder la question financière. Encore faut-il que *« le réalisme visuel soit en accord avec le niveau de communication qui s'établit entre l'agent virtuel et l'être humain »*.

Michel Alberganti