

science Le Monde & médecine



Les métamorphoses de la voix

Parlée ou chantée, la voix humaine se prête à d'infinies variations. Des informaticiens, des acousticiens et des chercheurs en neurosciences dissèquent et distordent en tous sens les paramètres vocaux... pour synthétiser des voix « sur mesure ». Au risque aussi de possibles détournements

FLORENCE ROSIER

Une folle arnaque a récemment défrayé la chronique. En septembre, le *Wall Street Journal* révélait qu'en mars des aigrefins avaient ravi la coquette somme de 220 000 euros à une entreprise britannique. Comment ? En simulant la voix du dirigeant de la maison mère allemande. La victime britannique avait reçu un appel téléphonique de son – faux – patron, lui ordonnant d'effectuer en urgence un virement sur un compte hongrois. Cette voix artificielle imitait

l'accent germanique du patron et ses intonations. « Le résultat n'est pas parfait, mais peut faire illusion dans un environnement sonore perturbé », confiait au *Monde* Jean-Baptiste Mounier, salarié d'Euler Hermes, l'assureur qui a divulgué l'affaire au *Wall Street Journal*. Les escrocs ont-ils eu recours à une intelligence artificielle (IA) pour créer une voix de synthèse ? Ont-ils fait appel à des procédures de *deep learning* (« apprentissage profond »), à partir d'un discours enregistré dudit patron ? Les experts n'y croient guère, notamment parmi ceux qui ont fait profession de manipuler en tous sens la voix humaine.

Place Igor-Stravinsky, à Paris. La clé de sol veille, flanquée de la monumentale Nana et des grands automates de métal. Baigné par l'ombre tutélaire du Centre Pompidou, l'Oiseau de feu les contemple. L'Amour, de sa bouche écarlate, nous invite à plonger en sous-sol.

L'Ircam, ce centre unique fondé par Boulez
Nous voici donc sous la fameuse fontaine de Jean Tinguely et Niki de Saint Phalle. Nous pénétrons dans les laboratoires de l'Ircam (Institut de recherche et de coordination acoustique-musique). Là, tout un ballet de musiciens-chercheurs s'active. Sous terre, pour limiter tout bruit parasite. C'est que cet institut, fondé par Pierre Boulez en 1970, est consacré à la création musicale contemporaine. Mais aussi au décodage, à la transformation et à la synthèse de la voix parlée ou chantée. Un centre unique, par sa dimension et son rayonnement, qui mêle art, science et technologies.

Transformer à façon la voix humaine ? La transfigurer, voire la contrefaire ? Voilà qui a de

quoi troubler, tant notre voix nous semble une part intime et indissociable de notre identité. Tant chaque voix nous semble singulière. Mais alors, l'arnaque à la voix truquée, qui s'ajoute à d'autres escroqueries vocales ? « Nous avons de forts doutes sur l'utilisation d'une intelligence artificielle dans ce contexte », explique Axel Roebel, responsable de l'équipe Analyse et synthèse des sons de l'Ircam (CNRS-Sorbonne université). Pour le moment, les agents conversationnels – l'Alexa d'Amazon, par exemple – sont capables de mener un dialogue mais ils manquent beaucoup d'expressivité. Avec de tels agents, imiter l'intonation et l'accent allemand d'une personne, dans un contexte de dialogue ouvert, me semble aujourd'hui impossible. »

Selon Alex Roebel, le *fake* a plutôt été produit par un « bon imitateur » qui a simulé l'accent et l'intonation du dirigeant. Puis les escrocs auraient utilisé un filtre pour s'approcher de la couleur de sa voix. Ce type d'escroquerie serait « à la portée de presque tout le monde » !

→ LIRE LA SUITE PAGES 4-5

La luminothérapie, aussi efficace que les antidépresseurs ?

Reconnue pour ses bienfaits contre la dépression saisonnière, la lumière blanche serait une thérapie viable pour toutes les formes dépressives, selon une méta-analyse française

PAGE 2



Alerte « rougeole » sur le système immunitaire

Le virus morbillieux réinitialise les défenses de l'organisme et le rend plus vulnérable à d'autres infections. Un « paradoxe immunitaire » que prévient la vaccination

PAGE 3



Portrait L'inlassable quête des origines de la vie

La géologue Frances Westall, qui dirige le groupe d'exobiologie du Centre de biologie moléculaire du CNRS à Orléans, se décrit elle-même comme une « vagabonde scientifique »

PAGE 8

Des voix déchiffrées, transformées, contrefaites...

► SUITE DE LA PREMIÈRE PAGE

Confortant cette analyse, une étude s'était intéressée en 2018 aux capacités des technologies informatiques actuelles à usurper une identité vocale, à partir de données éparpillées récoltées sur le Net (Jaime Lorenzo-Trueba *et al.*, «Speaker Odyssey, The Speaker and Language Recognition Workshop», p. 240 – 247). Son verdict : ces voix falsifiées sont de très mauvaise qualité. «Elles seraient bien en peine de duper quiconque!», conclut Nicolas Obin, enseignant-chercheur au laboratoire des Sciences et technologies de la musique et du son de l'Ircam (CNRS-Sorbonne université).

Malgré ces difficultés, une vague de start-up déferle aujourd'hui dans le domaine des technologies vocales. Qu'en est-il des performances de synthèse vocale annoncées à grands cris par des sociétés comme Lyrebird, au Canada? Dès 2017, Lyrebird se vantait d'avoir développé une technologie d'IA capable d'imiter n'importe quelle voix à partir d'un enregistrement d'«une minute seulement». «Leur but était de produire une phrase fixe, prononcée avec une voix de synthèse», précise Axel Roebel. On est loin d'un dialogue en temps réel.

«Il faut rester critique envers les annonces spectaculaires qui se nourrissent des fantasmes autour de l'IA. La réalité est bien plus mitigée, tempère Nicolas Obin. Aucune entreprise à ce jour n'est capable de cloner votre voix de manière réaliste et crédible, à partir de quelques enregistrements.» Il se peut qu'un jour cette prouesse devienne réalité, admet-il, mais «il reste bien du chemin à parcourir». Toujours est-il qu'un minimum semble ici requis : «disposer d'au moins 25 heures d'enregistrements de qualité de la voix à imiter.»

Il existe pourtant un fameux exemple de voix de synthèse très réaliste : c'est la voix standard de Google. Le géant propose aux Américains, depuis début 2019, un agent artificiel qui réserve pour eux une table au restaurant. «Cet exemple a probablement été très travaillé pour la démonstration», observe Nicolas Obin. Quant à l'autre service annoncé par Google – prendre un rendez-vous chez un coiffeur –, il n'a pas encore été déployé. «Cela montre la dépendance du système au contexte. C'est une de ses limites», note Axel Roebel.

Pour modifier à façon la voix humaine, la stratégie de l'Ircam diffère radicalement. Dans l'approche précédente, il s'agissait de créer, à partir de n'importe quel texte, une voix qui parle de manière naturelle. Avec les limites que l'on sait. A l'Ircam, les chercheurs ont donc rusé : au lieu de créer une voix *ex nihilo*, ils font d'abord appel à un comédien. «La voix de cet acteur nous sert de trame vocale, un peu à la manière de la "capture de mouvement" pour l'image», explique Nicolas Obin. Le comédien prononce, avec l'intonation souhaitée, les textes à reproduire. Ensuite, un algorithme transforme le timbre de sa voix pour «coller» au mieux à celui de la personne à imiter.

Cette ingénieuse approche a déjà permis aux chercheurs de reproduire les voix de Marilyn Monroe, Philippe Pétain ou Louis de Funès. Il leur a fallu seulement dix minutes d'enregistrements de la voix de ces personnalités. «Nous parvenons ainsi à un rendu d'une qualité acceptable par des professionnels du son», précise Nicolas Obin. Nous verrons plus loin comment.

Notre cerveau, un as du décodage vocal

Mais, auparavant, place à ces prodiges : nos propres systèmes d'analyse de la voix humaine! «Elle contient (...) un nombre étonnant de signaux : dès la première phrase au téléphone, on sait à qui l'on a affaire, son sexe, son âge, sa culture, son humeur agressive, abattue ou érotique, et même son niveau social», note Boris Cyrulnik dans *Les Nourritures affectives* (Odile Jacob, 1993). Que nous soyons capables de saisir, grâce aux réseaux de neurones de notre cerveau, les moindres nuances, les plus infimes vibrations d'une voix humaine : voilà qui donne la mesure du défi que pose l'imitation d'une telle voix, dans sa singularité!

En un éclair, nous percevons par exemple si l'inconnu qui nous parle est bienveillant ou malveillant, dominant ou dominé, digne de confiance ou pas. Par quel «miracle»? A l'Ircam, cette énigme a été en partie déchiffrée. «Nous avons conçu des algorithmes qui transforment et manipulent les différentes dimensions de la voix. Cela nous permet d'explorer la façon dont notre cerveau extrait, calcule et décode les signaux sociaux contenus dans la voix d'autrui», explique Jean-Julien Aucouturier, chercheur en sciences cognitives au CNRS.

C'est un terme banal qui a révélé le fin mot de cette histoire. Les chercheurs ont produit, à l'aide d'un algorithme, une infinité d'intonations du

mot «bonjour». Puis ils ont soumis ces variantes, par paires, à de nombreux volontaires. Avec cette question : lequel de ces deux «bonjour» vous semble le plus digne de confiance (ou le plus dominant...)? Résultat : «L'immense majorité des gens partagent la même représentation mentale de ce qu'est un "bonjour" digne de confiance ou un "bonjour" dominant...», résume Jean-Julien

EN UN ÉCLAIR,
NOUS PERCEVONS
SI L'INCONNU QUI NOUS
PARLE EST BIENVEILLANT
OU MALVEILLANT,
DOMINANT OU DOMINÉ



Aucouturier. Un «bonjour» digne de confiance, par exemple, est plutôt stable sur la première syllabe ; il monte légèrement sur la deuxième puis abruptement sur la fin. Un «bonjour» de chef, lui, descend sur la première syllabe puis encore plus sur la seconde. Une découverte publiée en mai 2018 dans les *PNAS*. Autrement dit, «les gens usent tous de la même partition mentale pour déchiffrer les signaux sociaux de la voix.» Mieux, ce code est généralisable à d'autres mots.

Certains patients victimes d'un accident vasculaire cérébral (AVC) ne parviennent plus à décoder les signaux sociaux dans la voix. Ils n'arrivent plus à reconnaître, notamment, une phrase interrogative – ni, en conséquence, à comprendre l'ironie. Pour autant, leurs capacités de compréhension du langage (syntaxe et sens des mots) sont intactes. «Ce n'est pas le contenu de la phrase qui pose problème, mais la façon dont elle est prononcée. Nous cherchons à déchiffrer l'algorithme cérébral responsable de ces erreurs, ce qui pourrait livrer des pistes de rééducation», espère Jean-Julien Aucouturier.

Un lion avec la voix de Sean Connery

A l'Ircam, le mystère du «sourire de la voix» a aussi été dissipé. Comment percevons-nous, au téléphone ou à la radio, si notre interlocuteur sourit ou non – sans même le voir? Quand il sourit, ses muscles zygomatiques étirent les coins de sa bouche, ce qui modifie les résonateurs de sa cavité buccale. «Un petit changement de timbre peut ici avoir un effet très puissant», note Jean-Julien Aucouturier. Son équipe a développé un logiciel qui simule l'effet du sourire sur la parole. Breveté, ce logiciel est utilisé dans la production musicale et les jeux vidéo. Il intéresse aussi les dispositifs médicaux, pour «humaniser» des synthétiseurs vocaux. «Avec ce logiciel, nous avons confirmé que les auditeurs reconnaissent les sourires dans la voix. Et montré qu'ils les imitent en souriant en retour, inconsciemment.» Des sourires si discrets qu'il faut, pour les révéler, des électrodes qui mesurent l'activité électrique des muscles du visage. L'étude a été publiée dans *Current Biology* en 2018. Mieux : l'équipe vient de reproduire ce travail chez des aveugles de naissance. Eux aussi, quand ils entendent une voix sourire, sourient en retour, alors même qu'ils n'ont jamais vu de sourire. Un écho à la *Lettre sur les aveugles à l'usage de ceux qui voient* (1749) où Diderot s'interroge sur les adaptations sensorielles des non-voyants.

De fait, tous les éléments de la chaîne de la voix peuvent traduire notre état psychique et émotionnel. En cas de stress, par exemple, notre voix tremble. C'est parce que le nerf vague, qui

UN VOYAGE SONORE, CHANTÉ ET ENCHANTÉ

C'est une expérience musicale unique et onirique. Un rêve éveillé où l'on chemine sous une coupole sonore. Pardessus nos têtes, une centaine de haut-parleurs diffusent des «nuages de voix» chantées ou chuchotées. On y reconnaît des airs familiers, des bribes de mélodies d'enfance. Et l'on y croise tour à tour un clarinetiste, un bassoniste, un clown ou une ballerine, comme sortis d'un songe.

L'expérience Lullaby («berceuse», en anglais) s'est donnée en juin au Centquatre à Paris, dans le cadre du festival ManiFeste de l'Ircam. La première mondiale avait eu lieu à Francfort (Allemagne), en février, avec l'Ensemble Modern de Francfort et des comédiens dirigés par le metteur en scène Claus Guth. «Il s'agit d'une expérience immersive, à la croisée des mondes du concert, de l'installation, du théâtre et de l'opéra», explique Jérôme Nika, chercheur et réalisateur en informatique musicale, qui a œuvré

avec Thierry Coduys à ce projet, conçu par le compositeur Pascal Dusapin.

Lullaby Experience illustre le pouvoir universel de la voix, quand elle renvoie à nos plus tendres années. Elle part «d'un matériau minimaliste et intime : la berceuse», raconte Jérôme Nika. Ou plutôt, des berceuses du monde entier. Mi-2018, un appel était lancé par l'Ircam, le Centre Pompidou et l'Ensemble Modern de Francfort : «Enfants comme adultes, participez à la création musicale en partageant en ligne vos mélodies d'enfance...» Les enregistrements collectés devaient fournir la «carte sonore» pour construire «une cartographie chantante du monde».

Des berceuses du monde entier

Plusieurs centaines de mélodies ont été collectées. La plupart venaient de France et d'Allemagne. Mais bien d'autres pays – d'Europe, d'Amérique, du Moyen-Orient et d'Asie – ont apporté leur voix à l'édifice. Seules

l'Afrique et l'Australie manquent encore à l'appel. Mais cette œuvre circulera : elle devrait intégrer des chansons de tous les continents. «Plus la variété des langues, des timbres de voix, des types de mélodies sera grande, plus l'œuvre finale sera riche», relève Jérôme Nika.

Les auteurs ont eu recours à un modèle informatique conçu par Jérôme Nika, avec l'équipe Représentations musicales de l'Ircam. «Les mélodies collectées ont servi de terrain d'apprentissage pour ce modèle. Mais aussi de vivier pour qu'il y puise des syllabes, des voix, des séquences mélodiques.» Dans un premier temps, ce modèle a «appris» ces chansons pour en comprendre la structure. Ensuite, il a produit de courtes polyphonies, en réponse à des intentions données. «Par exemple, nous lui demandions de créer une vaste polyphonie harmonique autour de L'Alouette est sur la branche, en ne puisant que dans des voix de femmes assez graves. Ou en-

core, nous lui commandions des pièces entièrement chuchotées.» Ce modèle avait un certain degré de liberté créative. «Nous avons été surpris par certains résultats!», confie Jérôme Nika. A partir des milliers de «briques» musicales obtenues, parfois retaillées, Pascal Dusapin a construit une forme musicale complète.

Encore fallait-il que le modèle comprenne les instructions. Et qu'il détecte les rythmes, identifie les mélodies, distingue les voix dynamiques ou monotones, claires ou voilées, chantées ou chuchotées... A l'Ircam, les chercheurs Axel Roebel et Nicolas Obin ont développé des algorithmes qui découpent ces chansons douces, syllabe par syllabe. Puis ils ont analysé le profil sonore de ces fragments, avant d'en effectuer un minutieux encodage. Les chanceux ayant cheminé dans ces nuages chantés – et enchantés – auront vécu un drôle de retour en enfance. ■

FL. R.



DE SI DÉLICATES EXPERTISES VOCALES

A l'université d'Avignon, Jean-François Bonastre, professeur d'informatique, s'intéresse aux systèmes de reconnaissance automatique de la voix. « Avec des laboratoires de police et de gendarmerie, nous travaillons à certifier les systèmes utilisés dans des affaires criminelles. » Le chercheur pointe leurs biais possibles. Par exemple, les systèmes commerciaux conçus pour des opérations bancaires ne sont pas adaptés au monde de la justice.

« Les particularités de la langue maternelle peuvent laisser des marques qui faussent l'expertise vocale », relève-t-il. Une fameuse affaire criminelle illustre ce type de failles. Jérôme Prieto, « le Basque condamné au son de sa voix », avait été accusé d'avoir jeté des cocktails Molotov sur des voitures de policiers, le 5 septembre 1996, à Hendaye. Sur la foi de deux expertises vocales (l'attentat avait été revendiqué par téléphone), ce jeune militant nationaliste a d'abord été condamné à trois ans de prison. Mais, en 1999, il fut relaxé en appel : les expertises vocales ont été jugées défaillantes. « Les Basques ont une façon particulière de prononcer certaines occlusives, ce qui a pu induire les experts en erreur », explique Jean-François Bonastre.

Boîtes noires

Son laboratoire cherche à comprendre comment fonctionnent les systèmes de reconnaissance automatique de la voix, ces boîtes noires. « Ils se fondent surtout sur l'analyse des voyelles, plus stables que les consonnes. Parmi elles, les plus fiables sont les voyelles nasales (on, en, in). Les consonnes nasales (m, n) sont aussi très utiles. » Non sans logique : la fosse nasale est une cavité stable, contrairement aux articulateurs de la cavité buccale, très mobiles. Mais si le locuteur est enrhumé, la technique est évidemment beaucoup moins fiable.

Et les systèmes de retranscription écrite de la parole ? « Depuis cinq ans, ils ont fait de grands progrès. Leur performance est aujourd'hui satisfaisante, même pour retranscrire la voix de différents locuteurs. » Reste qu'ils sont moins fiables pour les voix de femmes, pour les niveaux de langage familier, ou lorsque la voix est enregistrée dans des espaces bruyants. ■

FL. R.

EN PIOCHANT DES PHONÈMES APPROPRIÉS, LES CHERCHEURS ONT RÉUSSI À FAIRE CHANTER... MARGARET THATCHER. ÉTONNANTE EXPÉRIENCE

innerve notre larynx, agit sur les muscles qui modulent la résistance des cordes vocales.

Venons-en maintenant à cette troublante mission des chercheurs : imiter l'inimitable, au fond, c'est-à-dire une voix donnée, dans sa singularité et sa charge émotive. A l'Ircam, les chercheurs dissèquent les paramètres de la voix humaine... pour les distordre à loisir, selon l'effet recherché. Le monde du cinéma les sollicite régulièrement. Ils ont ainsi contribué à recréer la voix du mythique castrat italien du XVIII^e siècle Farinelli, pour les besoins du film de Gérard Corbiau (1994). « Son intonation était pure, son trille magnifique, son contrôle de la respiration extraordinaire et sa gorge très agile... », décrit son contemporain, le compositeur baroque Johann Joachim Quantz. Sa voix de cristal aurait même suscité des orgasmes !

Comment reconstruire cette voix de légende ? Farinelli pouvait couvrir trois octaves. Le directeur artistique du film a d'abord enregistré séparément un contre-ténor et une soprano colorature. Puis il a mixé les deux voix. Enfin, l'Ircam a unifié les deux timbres. Pour un résultat, au final, assez diversement apprécié.

Pour les besoins du film *Juger Pétain* (2014), de Philippe Saada, l'équipe d'Axel Roebel a reconstitué la voix parlée du Maréchal déchu, durant son procès. « L'enregistrement audio de ce procès avait été interdit, mais des vidéos avaient été tournées. On disposait aussi de sa retranscription écrite. Ainsi que de plusieurs enregistrements de la voix de Pétain, lors d'autres occasions », raconte Axel Roebel. Un acteur a d'abord prononcé le texte de la déclaration de Pétain, en imitant sa prosodie. « Puis nous avons appliqué à cette voix le timbre vocal de Pétain, tel que nous l'avions caractérisé. » Enfin, la production a fait coïncider le tout aux mouvements des lèvres de l'accusé, puis rajouté les échos de la salle.

Autre défi : transformer à loisir une voix enregistrée. Pour cela, les chercheurs ont développé une panoplie d'outils : traitement du signal, analyse statistique, apprentissage machine. Tel cet algorithme qui colore la voix d'émotions variées. « Ce logiciel transforme une voix neutre en une voix exprimant la colère, la tristesse ou la joie. Sa limite actuelle, c'est qu'il ne fonctionne que pour une voix donnée », explique Axel Roebel.

Avant de transformer une voix sur mesure, les chercheurs doivent d'abord démêler ses deux grandes composantes. Autrement dit, ils séparent la « source » de la voix (les vibrations des cordes vocales) de son « filtre » (leurs résonances dans le tractus vocal). Ensuite seulement, ils peuvent les manipuler séparément. Pour changer la hauteur de la voix, par exemple, ils ne manipulent que la

source (en réalité, c'est plus compliqué). Et pour modifier les paroles sans toucher au timbre, ils ne manipulent que le filtre. Illustration : les chercheurs ont ainsi fait parler un lion... avec la voix de Sean Connery. L'astuce a consisté à caractériser la source vocale du félin, puis à insérer cette source dans le filtre de la voix de l'acteur. Et celui-ci de parler en rugissant, avec conviction.

La voix de synthèse est un vieux défi. Présente dans nos GPS et nos smartphones, elle fait partie de notre quotidien. Ainsi de l'Alexa d'Amazon, qui imite assez bien la voix humaine. « Mais ces voix artificielles sont souvent monotones. A l'aide de réseaux d'apprentissage profond, nous essayons de les rendre plus expressives », indique Axel Roebel. Pour autant, avec de tels outils, « nos résultats de conversion d'identité vocale restent pour le moment assez limités », admet le chercheur. Ces limites, ajoute-t-il, touchent tous les laboratoires qui travaillent sur cette question. « De plus, ces systèmes ne fonctionnent pas en temps réel : ils ne permettent pas un vrai dialogue. »

Trois siècles de recherche sur la voix

Traduire nos pensées en une voix de synthèse, maintenant. Pur fantasme ? Peut-être pas. Un premier pas dans ce sens a été accompli par l'équipe de Gopala Anumanchipalli, à l'université de Californie. L'exploit a été relaté dans la revue *Nature* du 24 avril. Cinq volontaires ont été recrutés : des patients épileptiques dont la prise en charge nécessitait l'implantation d'électrodes dans le cerveau. Les chercheurs leur ont demandé de lire des centaines de phrases à voix haute. En parallèle, les signaux électriques émis par leurs neurones, dans les aires qui contrôlent l'articulation, étaient enregistrés. Ensuite, un logiciel d'IA a couplé ces signaux neuronaux aux paroles prononcées. Puis il a construit un « tractus vocal virtuel » (larynx, langue, lèvres...), dont les mouvements reproduisaient ceux du « vrai » tractus du patient en train de parler. Dernière étape : convertir l'activité du tractus virtuel en signal audio – c'est-à-dire en voix parlée. Sur

101 phrases ainsi synthétisées, 43 % ont été pleinement intelligibles. Pour les auteurs, la faisabilité de cette approche est démontrée. A terme, un tel décodeur pourrait aider ceux qui ont perdu la parole à communiquer. Mais la méthode reste très invasive.

Une brève incursion dans le passé permet toutefois de mesurer le chemin accompli. Les voix de synthèse sont le fruit de trois siècles de recherches. En 1791, une machine à parler mécanique (« Speaking machine ») a été inventée par le Hongrois Wolfgang von Kempelen. Il faudra attendre 1939 pour qu'un autre palier soit franchi. Aux Etats-Unis, les laboratoires Bell conçoivent le premier synthétiseur vocal électronique, le Voder, suivi dans la foulée du Vocoder (« Voice Encoder »). Par l'intermédiaire d'un clavier, la machine obéissait aux commandes d'un opérateur, tandis que ses effets sonores étaient modulés par des pédales. Reconnaisable entre tous, son timbre métallique inspirera les voix de nombreux robots de cinéma, tel le C-3PO de *Star Wars*.

Les années 1960, ensuite, ont vu naître les premières voix de synthèse informatique, tel l'IBM 704. Deux à trois décennies plus tard, un bond qualitatif sera franchi, grâce à la démultiplication des capacités de stockage et de traitement des ordinateurs... Bel exploit : en 1985, le fameux air de *La Reine de la nuit* a été reconstitué par un logiciel de synthèse, à l'Ircam, par l'équipe de Xavier Rodet. Mozart à l'ère informatique. Mais il s'agissait de vocalises, non de paroles articulées. Depuis lors, le projet ChanTeR s'est attaché à synthétiser du chant expressif, à partir d'un texte et d'une mélodie. Pour mener à bien ce projet, les chercheurs ont construit une base de données couvrant tous les diphtonges (des séquences de deux phonèmes) chantés de la langue française. Pour cela, ils ont enregistré deux chanteurs professionnels, un homme et une femme. Ensuite, il leur a suffi de « piocher dans cette base les phonèmes appropriés. Puis de les mettre bout à bout et de les transposer avec les bonnes hauteurs et les bonnes intensités, à l'aide de logiciels », explique Axel Roebel. Et la parole devient chant, avec la mélodie et l'accent voulu. Les chercheurs ont ainsi fait chanter... Margaret Thatcher. Étonnante expérience.

Autre démonstration. « C'est une chanson qui nous ressemble... », entonne Edith Piaf, de sa voix sans pareille. Le logiciel Isis tente bravement de l'imiter. Pas facile... S'il intègre un modèle de « style de voix » de l'artiste, le résultat est plus convaincant. Mais il manque encore à cet avatar vocal un rien... d'inimitable, précisément. ■

FLORENCE ROSIER