



COMMUNIQUÉ DE PRESSE NATIONAL | PARIS | 13 OCTOBRE 2017

Une échographie linguale augmentée pour la rééducation orthophonique

Une équipe de chercheurs du laboratoire GIPSA-lab (CNRS/Université Grenoble Alpes/Grenoble INP) et d'Inria Grenoble Rhône-Alpes vient de mettre au point un système permettant de visualiser, en temps réel, nos propres mouvements de langue. Capturés à l'aide d'une sonde échographique placée sous la mâchoire, ces mouvements sont traités par un algorithme d'apprentissage automatique qui permet de piloter une « tête parlante articulatoire ». En plus du visage et des lèvres, cet avatar fait apparaître la langue, le palais et les dents habituellement cachés à l'intérieur de l'appareil vocal. Ce système de « retour visuel », qui devrait permettre de mieux comprendre et donc de mieux corriger sa prononciation, pourra servir à la rééducation orthophonique ou l'apprentissage d'une langue étrangère. Ces travaux sont publiés dans la revue *Speech Communication* d'octobre 2017.

La rééducation orthophonique d'une personne atteinte d'un trouble de l'articulation s'appuie en partie sur la répétition d'exercices : le praticien analyse qualitativement les prononciations du patient et lui explique oralement ou à l'aide de schémas comment placer ses articulateurs, et notamment sa langue, dont il n'a généralement que peu conscience. L'efficacité de la rééducation repose donc sur la bonne intégration par le patient des indications qui lui sont données. C'est à ce stade que peuvent intervenir les systèmes de « retour articulatoire visuel » qui permettent au patient de visualiser en temps réel ses propres mouvements articulatoires (et notamment les mouvements de sa langue) afin de mieux en prendre conscience et donc de corriger plus rapidement ses défauts de prononciation.

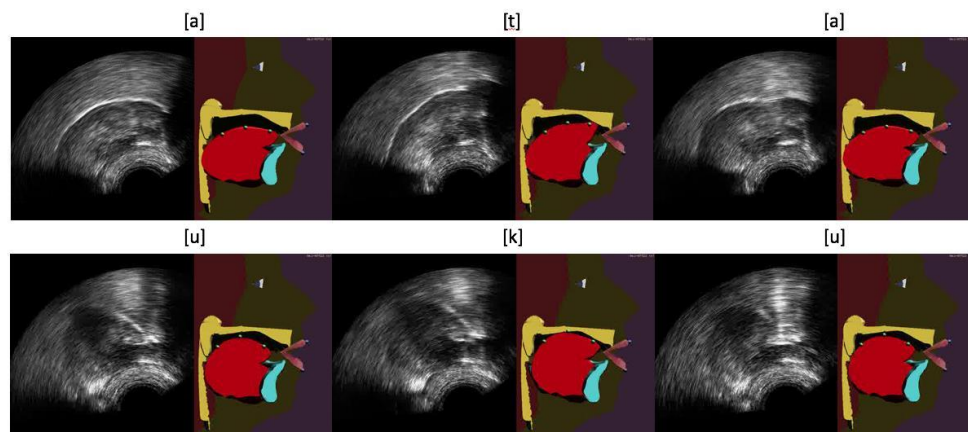
Depuis quelques années, des chercheurs anglo-saxons semblent privilégier la technique de l'échographie pour la conception de ces systèmes de retour visuel. L'image de la langue est alors obtenue en plaçant sous la mâchoire d'un locuteur une sonde analogue à celle classiquement utilisée pour l'imagerie du cœur ou du fœtus. Cette image est parfois jugée difficile à exploiter par le patient car elle n'est pas de très bonne qualité et ne donne aucune information sur la place du palais et des dents. Dans ces nouveaux travaux, les chercheurs français proposent d'améliorer ce retour visuel en animant automatiquement et en temps réel une tête parlante articulatoire à partir des images échographiques. Ce clone virtuel d'un véritable locuteur, en développement depuis de nombreuses années au GIPSA-lab, permet une visualisation plus intuitive, car contextualisée, des mouvements articulatoires.

La force de ce nouveau système repose sur un algorithme d'apprentissage automatique (*machine learning*), sur lequel les chercheurs travaillent depuis plusieurs années. Cet algorithme permet (dans une certaine limite) de traiter des mouvements articulatoires que l'utilisateur ne maîtrise pas encore lorsqu'il commence à utiliser le système. Cette propriété est indispensable pour les applications thérapeutiques



visées. Pour atteindre une telle performance, l'algorithme exploite un modèle probabiliste construit à partir d'une grande base de données articulatoires acquises sur un locuteur dit « expert », capable de prononcer l'ensemble des sons d'une ou plusieurs langues. Ce modèle est adapté automatiquement à la morphologie de chaque nouvel utilisateur, lors d'une courte phase d'étalonnage du système, au cours de laquelle le patient doit prononcer quelques phrases.

Ce système, validé en laboratoire pour des locuteurs sains, est aujourd'hui testé dans une version simplifiée dans une étude clinique pour des patients ayant subi une intervention chirurgicale de la langue. Par ailleurs, les chercheurs développent aussi une autre version du système dans laquelle la tête parlante articulatoire est animée automatiquement, non pas à partir d'échographie, mais directement à partir de la voix de l'utilisateur¹.



Exemple d'animations du modèle de langue de la tête parlante articulatoire du GIPSA-lab à partir d'images échographiques, à l'aide de l'algorithme « Integrated Cascaded Gaussian Mixture Regression » pour les séquences [ata] (haut) et [uku] (bas). © Thomas Hueber / GIPSA-Lab (CNRS/Université Grenoble Alpes / Grenoble INP).

Une vidéo est également disponible sur : <https://www.youtube.com/watch?v=u8jb4b0fMsE>

Bibliographie

Automatic animation of an articulatory tongue model from ultrasound images of the vocal tract. Fabre, D., Hueber, T., Girin, L., Alameda-Pineda, X., Badin, P. (2017). *Speech Communication*, vol. 93, pp. 63-75.

¹ Voir **Speaker-Adaptive Acoustic-Articulatory Inversion using Cascaded Gaussian Mixture Regression**. Hueber, T., Girin, L., Alameda-Pineda, X., Bailly, G. (2015), in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2246-2259.



www.cnrs.fr



Contacts

Chercheur CNRS | Thomas Hueber | T 04 76 57 49 40 | thomas.hueber@gipsa-lab.fr

Presse CNRS | Alexiane Agullo | T 01 44 96 43 90 | alexiane.agullo@cnrs-dir.fr