



PRESS RELEASE | PARIS | 13 OCTOBER 2017

Augmented tongue ultrasound for speech therapy

A team of researchers in the GIPSA-Lab (CNRS/Université Grenoble Alpes/Grenoble INP) and at INRIA Grenoble Rhône-Alpes has developed a system that can display the movements of our own tongues in real time. Captured using an ultrasound probe placed under the jaw, these movements are processed by a machine learning algorithm that controls an “articulatory talking head.” As well as the face and lips, this avatar shows the tongue, palate and teeth, which are usually hidden inside the vocal tract. This “visual biofeedback” system, which ought to be easier to understand and therefore should produce better correction of pronunciation, could be used for speech therapy and for learning foreign languages. This work is published in the October 2017 issue of *Speech Communication*.

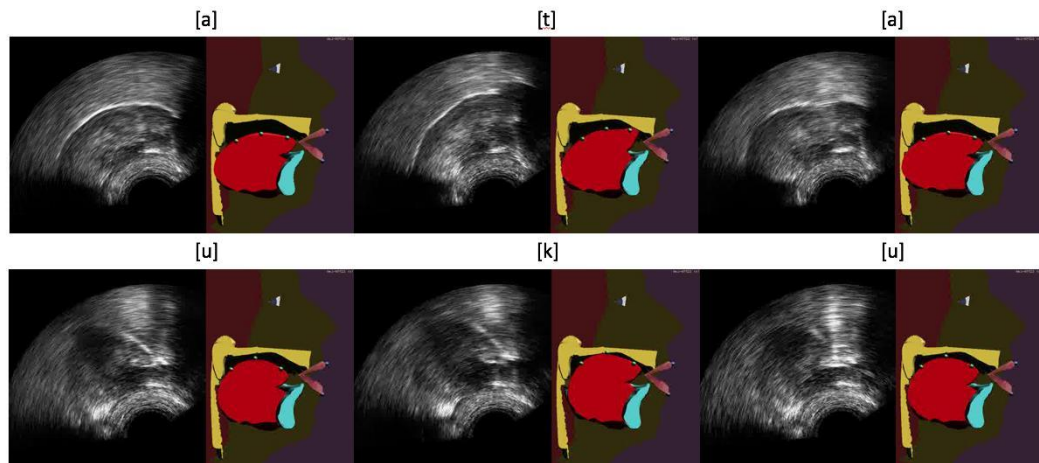
For a person with an articulation disorder, speech therapy partly uses repetition exercises: the practitioner qualitatively analyzes the patient’s pronunciations and orally explains, using drawings, how to place articulators, particularly the tongue: something patients are generally unaware of. How effective therapy is depends on how well the patient can integrate what they are told. It is at this stage that “visual biofeedback” systems can help. They let patients see their articulatory movements in real time, and in particular how their tongues move, so that they are aware of these movements and can correct pronunciation problems faster.

For several years, researchers have been using ultrasound to design biofeedback systems. The image of the tongue is obtained by placing under the jaw a probe similar to that used conventionally to look at a heart or fetus. This image is sometimes deemed to be difficult for a patient to use because it is not very good quality and does not provide any information on the location of the palate and teeth. In this new work, the present team of researchers propose to improve this visual feedback by automatically animating an articulatory talking head in real time from ultrasound images. This virtual clone of a real speaker, in development for many years at the GIPSA-Lab, produces a contextualized — and therefore more natural — visualization of articulatory movements.

The strength of this new system lies in a machine learning algorithm that researchers have been working on for several years. This algorithm can (within limits) process articulatory movements that users cannot achieve when they start to use the system. This property is indispensable for the targeted therapeutic applications. The algorithm exploits a probabilistic model based on a large articulatory database acquired from an “expert” speaker capable of pronouncing all of the sounds in one or more languages. This model is automatically adapted to the morphology of each new user, over the course of a short system calibration phase, during which the patient must pronounce a few phrases.



This system, validated in a laboratory for healthy speakers, is now being tested in a simplified version in a clinical trial for patients who have had tongue surgery. The researchers are also developing another version of the system, where the articulatory talking head is automatically animated, not by ultrasounds, but directly by the user's voice¹.



Example of tongue model animations of the GIPSA-Lab articulatory talking head from ultrasound images, using the Integrated Cascaded Gaussian Mixture Regression algorithm for [ata] (top) and [uku] (bottom) sequences. © Thomas Hueber / GIPSA-Lab (CNRS/Université Grenoble Alpes / Grenoble INP).

A video is also available at: <https://www.youtube.com/watch?v=u8jb4b0fMsE>

Bibliography

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

Contacts

CNRS Researcher | Thomas Hueber | T + 33 (0) 4 76 57 49 40 | thomas.hueber@gipsa-lab.fr
CNRS Press Officer | Alexiane Agullo | T +33 (0) 1 44 96 43 90 | alexiane.agullo@cnrs-dir.fr

¹ See **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.