# Assistive speech technologies based on ultrasound imaging

Thomas Hueber, PhD - CNRS researcher - GIPSA-lab, Grenoble

Within the last four decades, ultrasound imaging has been used in phonetics to visualize tongue movements during speech production. Ultrasound imaging is a non-invasive and clinical-safe technique providing vocal tract scans in the mid-sagittal or coronal plane, with good spatial and temporal resolution. More recently, ultrasound imaging has also been used in the context of assistive speech technologies, as an input sensor for devices dedicated to speech-impaired people. These devices aim either at supplementing a speech production system that is no longer functional, for instance after a total laryngectomy, or at helping people to improve their pronunciation by visualizing directly their own articulatory movements (visual biofeedback).

In the first part of the talk, I will start recalling briefly the basic physics of ultrasound, and some practical aspects of tongue ultrasound imaging. I will present different strategies to control the position of the probe relative to the skull, and I will present the *Ultraspeech* platform [1] that allows the synchronous and simultaneous acquisition of high-speed ultrasound, video, and audio signals. Finally, I will review some image processing techniques that can be used to parameterize the ultrasound image, including the PCA-based *EigenTongue* technique [2].

In the second part of the talk, I will present some recent work on a *silent speech interface (SSI*, [3]*)* that aims at synthetizing in real-time an intelligible speech signal, only from articulatory movements captured during "silent articulation" (*i.e.* the speaker articulates normally but does not vocalize the speech sound). SSIs may be used by laryngectomized patients, as an alternative to the tracheoesophageal voice, or by any other speaker who wants to communicate quietly, or in very noisy environments.

The last part of the talk will be dedicated to our recent work on the use of ultrasound imaging for pronunciation training and biofeedback. First, I will present the *Ultraspeech-player* software [4] that allows an intuitive visualization of ultrasound articulatory data. A preliminary evaluation of this tool will be presented in the context of the rehabilitation of the [tr]/[kr] substitution. Finally, I will present some recent development of our visual biofeedback system, in which ultrasound is used to drive the tongue model of the GIPSA-lab *articulatory talking head*.

[1] Hueber, T., Chollet, G., Denby, B., and Stone, M. (2008). "Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application," Proceedings of International Seminar on Speech Production (Strasbourg, France), pp. 365-369.

[2] Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., and Stone, M. (2007). "Eigentongue feature extraction for an ultrasound-based silent speech interface," Proceedings of ICASSP (Honolulu, USA), pp. 1245-1248.

[3] Hueber, T., Benaroya, E.L., Chollet, G., Denby, B., Dreyfus, G., Stone, M., (2010) "Development of a Silent Speech Interface Driven by Ultrasound and Optical Images of the Tongue and Lips", Speech Communication, 52(4), pp. 288-300.

[4] Hueber T., "Ultraspeech-player: Intuitive visualization of ultrasound articulatory data for speech therapy and pronunciation training" (2013), Proceedings of Interspeech (show&tell), Lyon, France, 2013, pp.752-753.

[5] Fabre D, Hueber T., Badin P. (2014), « Automatic animation of an articulatory tongue model from ultrasound images using Gaussian mixture regression », submitted to Interspeech.