Synchronized speech, tongue ultrasound and lip movement video recordings with the “Micro” system

Tamás Gábor Csapó[1,4], Andrea Deme[2,4], Tekla Etelka Gráczi[3,4],
Alexandra Markó[2,4] and Gergely Varjasi[2,4]
[1]Budapest University of Technology and Economics,
[2]Dept. of Phonetics, Eötvös Lorand University,
[3]Research Institute for Linguistics, HAS, Hungary
[4]MTA-ELTE “Momentum” Lingual Articulation Research Group

This demonstration will show the details of the articulatory investigations (tongue ultrasound and lip movement) of the MTA-ELTE “Momentum” Lingual Articulation Research Group, including the technical aspects, applied hardware and software elements, sample recordings, and current and planned research.

Analysis of the shapes and dynamics of the human tongue during speech is important for modeling the vocal tract. Ultrasound imaging of the tongue is an attractive solution because it images tongue motion at a rapid frame rate (up to 100 Hz), which can capture subtle and swift movements during speech production (Stone, 2005). Csapó et al. (2017) have demonstrated the type and quality of the first speech and ultrasound recordings with the “Micro” (previously SonoSpeech) system (Articulate Instruments Ltd.). In the current demonstration, we will extend this by showing how to record video of the lips in synchrony with the ultrasound and speech signals.

Five Hungarian subjects (three females and two males) with normal speaking abilities were recorded while reading aloud sentences and nonsense words. The tongue movement was recorded in midsagittal orientation using a “Micro” ultrasound system (Articulate Instruments Ltd.) with a 2-4 MHz / 64 element 20mm radius convex ultrasound transducer at 80-100 fps. During the recordings, the transducer was fixed using an ultrasound stabilization headset (Articulate Instruments Ltd.). The video of the lips of one speaker was recorded at 59.94 fps (interlaced) either from front or from side view with an NTSC microcamera that was attached to the helmet (see Fig. 1). The video was digitized using a DFG2USB device. The speech was recorded with an Audio-Technica - ATR 3350 omnidirectional condenser microphone that was clipped approximately 20cm from the lips. The ultrasound and the audio signals were synchronized applying the frame synchronization output of the equipment with the Articulate Assistant Advanced software (Articulate Instruments Ltd.). The lip video and the audio signals were synchronized using a SyncBrightUp Unit (Articulate Instruments Ltd.) which adds a white mark to several video frames at the same time as putting a trigger signal to the audio. Both the microphone signal and the ultrasound synchronization signals were digitized using an M-Audio – MTRACK PLUS external sound card at 22050 Hz sampling frequency.

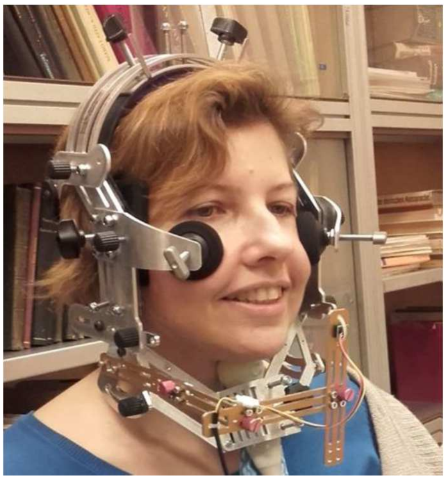

Figure 1: Ultrasound stabilization helmet with the fixed ultrasound transducer and lip camera (in the front of the lips)

After the recordings, the ultrasound frames were extracted as raw scan line data and converted to JPG images. Next, videos were constructed from the raw ultrasound data, lip movement and synchronized speech recordings (for sample images, see Fig. 2).

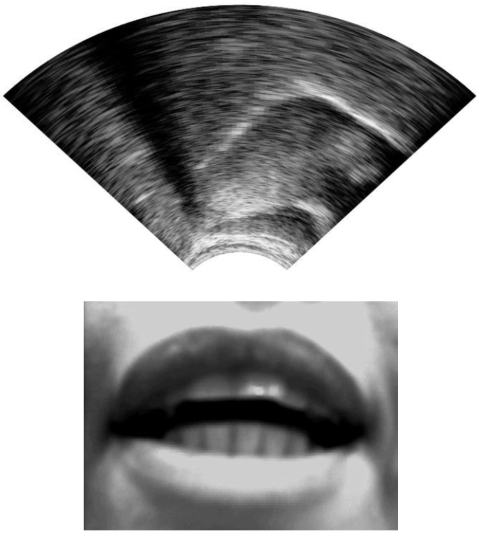

Figure 2: Sample ultrasound and lip images from a female speaker

The demonstration will include a general introduction of the hardware and software components of the “Micro” system, and sample tongue ultrasound and lip movement videos from the five speakers.

References

Csapó T. G., Deme A., Gráczi T. E., Markó A., Varjasi G., 2017. *Szinkronizált beszéd- és nyelvultrahang-felvételek a SonoSpeech rendszerrel* [Synchronized speech and ultrasound recordings with the SonoSpeech system], In: XIII. Magyar Számítógépes Nyelvészeti Konferencia [13th Conference on Hungarian Computational Linguistics] (MSZNY2017), Szeged, Hungary, 339–346.

Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. *Clin. Linguist. Phon.* 19, 455–501.