Evaluation of the uncertainty of multimodal articulatory data

Michael Aron, Marie-Odile Berger, Erwan Kerrien
LORIA/ INRIA Nancy-Grand Est, BP 101, 54602 Villers les Nancy, France
email: {aron,berger,kerrien}@loria.fr

Our purpose is to present the last improvements brought to our acquisition system of multimodal articulatory data and reports on a complete evaluation of the uncertainty attached to the data it provides.

In previous works [1, 2, 3], we presented our system to acquire multimodal articulatory data. The data are both dynamic: 2D Ultrasound (US) to get midsagittal images of the tongue at 66 Hz, electromagnetic (EM) sensors that provide their 3D position and orientation at 40 Hz, and audio data; and static: 3D Magnetic Resonance Imaging (MRI) depicting the vocal tract for various sustained articulations. The EM sensors are central in our system. Two 5DOF EM sensors are glued onto the tongue (apex and dorsum) to complete for possibly lacking US information. Besides, the US probe as well as the speaker’s head motions are tracked each by one 6DOF sensor. All the data are processed, synchronized and spatially registered [1] (figure 1).

The previous versions of our system implied quite a complex set of synchronization signals [2]. We introduced a single control PC whose hardware was carefully chosen so that it is used to record all dynamic data, but US, thus eliminating the need for most synchronization signals.

Besides the system improvements, we also performed a complete evaluation of the quantitative uncertainty attached to each data individually. A robotic table was used as reference to estimate the accuracy of the 5DOF EM sensors: 0.76 mm for the position and 0.5° on the angle of rotation are reported. An US phantom was used to calibrate the resolution of US images: the pixel size was found different from the manufacturer’s indication with a standard deviation of 0.007 mm/pixel. The uncertainty on the extraction of MRI data correspond to the voxel size which, in our acquisition protocol, is 0.625×0.625×3 mm³. All these measurements impact the rigid transformations $T$ which are involved in the spatial registration of data (figure 1).

We studied the global uncertainty of each transformation involved in the registration process, and especially the US to MRI registration (figure 2.a). Monte-Carlo statistical method [4] was used to estimate the uncertainty of this complex registration process. For each modality, a random noise, corresponding to the individual error computed above, was applied on each data used for calibration. A set of noisy transformations $T$ was thereafter computed. The Monte-Carlo method consists in computing a high number of registrations built from these noisy transformations. Four test points were identified in a typical US image of the tongue, and their images through the noisy registrations were computed, giving rise to four clouds of points in the MRI coordinate space. The RMS error found on these points is 2.4 mm for the US to MRI registration process (figure 2.b).
This study also enabled us to isolate the major sources of error in the registration process. Not surprisingly the EM sensors are an important factor, but the US resolution was also found to be critical, being responsible for 0.9 mm RMS error.

To our knowledge, this is the first multimodal system which provides a quantitative measurement about the uncertainty on articulatory data. This can be used to: (i) improve the computation of the transformations $T$ which most significantly impact the registration error; (ii) provide a basis for comparing acquisition systems.

We also think that the proposed framework might help the community to better understand the accuracy needed for articulatory data for speech applications.

![Figure 2: Fusion of US and MRI data. (a) 3D image of MRI, US and EM data. (b) Palate extracted from MRI data and superimposed on the US image. The red circle is the computed uncertainty.](image)

References


