Collaborative Quantitative Analysis of RT-MRI

N. Almeida[1], S. Silva[1], A. Teixeira[1], C. Cunha[2]
[1] Dep Electronics Telecom. & Informatics / IEETA, University of Aveiro, Portugal
[2] IPS, LMU Munich, Germany

1. Introduction

The study of the dynamic aspects of speech production can profit from real-time magnetic resonance (RT-MRI) to, for example, improve knowledge on which articulators and gestures are involved in producing specific sounds and foster improved speech production models. For a speech researcher non-specialist in image processing use RT-MRI data to test her/his hypothesis is not a simple task. This difficulty, adding to the costs, complex setups and the higher demands to subjects, prevents processing of data from a large number of subjects. The processing of RT-MRI acquisitions of many subjects at increasing frame rates (above 100Hz) is only possible by an automatization and use of data-driven approaches, to avoid the time-consuming annotation, errors, and inconsistencies associated with manual processes.
The community has contributed in recent years with data-driven approaches to extract and analyze features of interest such as contours and relevant landmarks (e.g. [1, 2]), but the use of such methods needs to be simplified. Speech production researchers need tools to allow complex, state-of-the-art processing and analyzes, as they have for acoustical analyses.

2. Proposal

We consider that processing of RT-MRI can be made accessible to non-specialists by tools that: (1) enable them to organize, annotate and process data from MRI (and other methods); (2) provide off-the-shelf processing components to create easily complex processing pipelines; (3) make possible collaboration and distribution of tasks by different researchers; (4) have ubiquitous access (anywhere, any computing platform).

3. Collaborative RT-MRI Processing Framework

The Architecture of the Tool/Framework implementing our proposal is presented in Figure 1. The Backend modules manages the interactions with users and serves as a gateway store uploaded data, to view information and to invoke the processing pipeline. The processing pipeline executes intensive operations in the users' data. Then, the next module analyses the result and saves the results in the database. The framework is cloud based, this way, allows users to access it at any time/anywhere remotely.

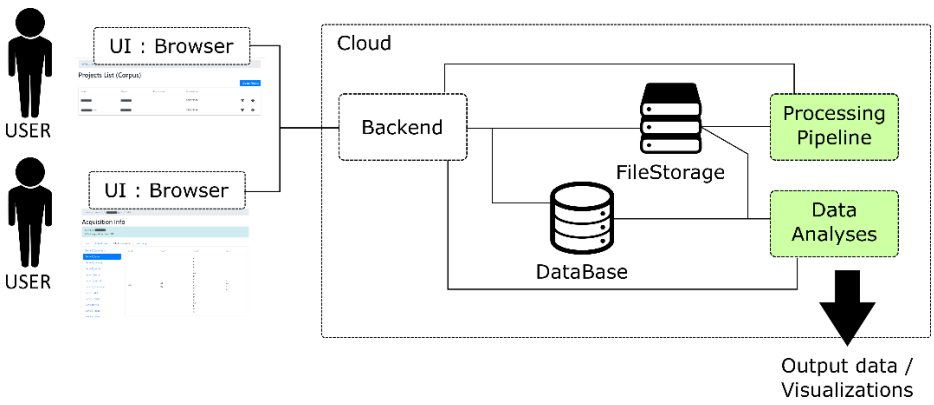

Figure 1 - General Architecture of the Framework for Collaborative Quantitative Processing of RT-MRI.

This tool allows an authenticated user to create a new project and upload each acquisition of that project. It has features to enable the segmentation of a small set of images: (1) the user selects the image; (2) marks each point (the landmark and auxiliary points) to define the contour (see Figure 2); in the end (3) the location can be adjusted and add or remove points; (4) finally it is selected if it is oral or nasal sound. The segmented images are processed to create a model to automatically segment the entire set of images. After the automatic segmentation, the user can revise the quality of it, correct manually the segmentation (correct the location of the points) and validate.

Figure 2 – Interface for the creation of initial segmentations.

Pipeline:

It constitutes the main processing component of the architecture. The vocal tract outlines are extracted adopting the method proposed by Silva et al. [3], resulting in contours identifying the different regions of interest. The comparison among vocal tract configurations was performed adopting and extending a previously proposed framework [3] enabling normalized quantification of differences, for different articulators/regions of the vocal tract and their visualization: velum (VEL), tongue dorsum (TD), tongue back (TB), tongue tip (TT), lip protrusion (LP), lip aperture (LA) and pharynx (Ph). Additionally, the work presented here also considered constrictions, characterized by location (CL) and sagittal distance (CD) at three regions [4].

3. Conclusion

The need for a new generation of tools to process increasing amounts of RT-MRI data by speech production researchers in general is defended and a Framework is proposed. Concrete examples of application of the framework in Phonetic research will be presented at the Symposium (if submission accepted). Examples will include the extension to a larger set of speakers of authors' recent study on the role of oral configurations in European Portuguese nasal vowels [4].

References

[1] C. Lammert, M. I. Proctor and S. S. Narayanan, "Data-driven analysis of realtime vocal tract MRI using correlated image regions," in InterSpeech, 2010.

[2] S. Silva and A. Teixeira, "Unsupervised segmentation of the vocal tract from real-time MRI sequences," Computer Speech and Language, p. 25–46, 2015.

[3] S. Silva and A. Teixeira, "Quantitative systematic analysis of vocal tract data," Computer Speech and Language, vol. 36, pp. 307-329, 2016.

[4] Cunha, S. Silva, A. Teixeira, C. Oliveira, P. Martins, A. Joseph and J. Frahm, "On the role of oral configurations in European Portuguese nasal vowels," in InterSpeech, Graz, 2019.

Acknowledgments

This work is partially funded by IEETA Research Unit funding (UID/CEC/00127/2019), by Portugal 2020 under the Competitiveness and Internationalization Operational Program, and the European Regional Development Fund through project SOCA – Smart Open Campus (CENTRO-01-0145-FEDER-000010), project MEMNON (POCI-01-0145-FEDER-028976) and by the German Federal Ministry of Education and Research (BMBF, with the project 'Synchonic variability and change in European Portuguese').