2019-01542 - PhD Position F/M Deep supervision of the vocal tract shape for articulatory synthesis of speech

Mission confiée

Context
The production of speech requires a signal source, i.e. the vibration of vocal folds, a noise of turbulence in the vocal tract, and a system of resonant cavities, i.e. the vocal tract. Speech articulators (jaw, tongue, lips, larynx, soft palate and epiglottis) are used to modify the shape of the vocal tract, and therefore the acoustic properties including the resonances of the vocal tract. When learning speech or a second language, speakers learn how to control the articulators to produce intelligible speech. Articulatory synthesis mimics this process by using the deformations of the vocal tract, and the parameters of vocal fold control over time. Thus, the interest of articulatory synthesis is to explain the articulatory origin of phonetic contrasts, change the position of articulators (or even block one of them), modify the control parameters of vocal folds, enable a realistic adaptation to a new speaker by modifying the size and shape of the articulators, and finally give access to physical quantities (e.g. pressure) in the vocal tract for example) without requiring the introduction of sensors in the vocal tract. Compared to other approaches to synthesis that offer a high level of quality, the strenght of articulatory synthesis is above all to control the entire process of speech production.

The generation of the geometric shape of the vocal tract at each point of the synthesis is most often based on the use of an articulatory model [1,2] that gives the shape of the tract with a small number of parameters. Each of the parameters corresponds to a deformation mode of the articulator considered, and the tongue being the most deformable articulator requires at least 6 parameters. The articulatory model is constructed from about 100 static MRI images of the vocal tract.

Principales activités

Project description
Recently we have been equipped with a two-dimensional real-time MRI acquisition system (at 55 images per second) for the vocal tract as part of a collaboration with the IADI laboratory (INSERM U1254) at Nancy hospital, and a database of several hours of speech for a speaker. The quality of these images of the mid-sagittal shape of the vocal tract is very good and it is therefore possible to track the contour of the articulators independently of the others because speech involves complex compensatory and coordinating gestures that would be lost if the vocal tract is processed in one piece[7].

The most important part of the work will be devoted to controlling the shape of the vocal tract. The idea is to develop a deep learning approach to determine the position of the articulators according to the phonemes to be articulated. The constraint is to be able to identify the role of each articulator in sufficient detail so as to be able to control its impact on the overall shape of the vocal tract, and to study coordination and compensation strategies between the articulators.

The abduction and adduction gestures of the vocal folds can be recorded using electro-photoglottography [8] and, as for the articulatory parameters, it will be possible to learn them according to the sequence of phonemes to be articulated. These two datasets will be fed into digital acoustic simulations [9] to verify the quality of the speech produced, and to study the articulatory factors of expressive speech.

References
Compétences

Required qualifications

MSc in computer science, signal processing or applied mathematics.

Language

French or English.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération


Monthly salary after taxes : around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).