Coordination of vocal folds activity with supraglottal cavities for synthesizing French consonants
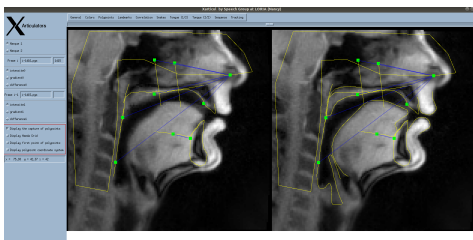
Ruoxiao YANG

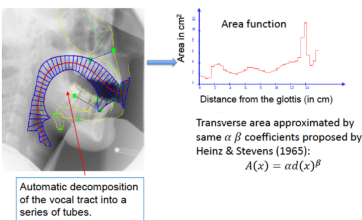Université de Lorraine

Background & Objective

- Articulatory copy synthesis models the natural speech production process by simulating the acoustic process of speech production instead of using a pre-recorded speech database.

- Two key parameters are controlled in articulatory copy synthesis: (1) Supraglottal cavities: by obtaining vocal tract (VT) shape through Magnetic Resonance Imaging (MRI) and (2) Vocal folds activity: by measurements of ElectroPhotoGlottoGraphy (EPGG).

- Why this approach: (1) Observe and study directly the physiology of human speech production and (2) Observe the link between articulatory gestures and acoustic domains during speech production.

- Recent progress in simulating the French fricative sounds (e.g., [s] in the French word "assis" and [z] in the word "Asie") in the framework of articulatory copy synthesis (Elie and Laprie, 2017), with incorporating the vocal folds opening data through the measurements of EPGG.

- The major **objective** of this project is to synthesize French stop consonants through articulatory copy synthesis.

Vocal tract (VT) shape and area functions

- The contours of articulators at the VT are extracted semi-automatically or by hand with the software package called "**Xarticulators**".

Recovering the area function

Transverse area approximated by same α β coefficients proposed by Heinz & Stevens (1965):
$$A(x) = \alpha d(x)^{\beta}$$

Automatic decomposition of the vocal tract into a series of tubes.

Vocal folds activity: measurements of EPGG

- Patterns of glottal opening for different segments:
  - The value of glottal opening area for consonants: unvoiced fricatives > voiced fricatives > unvoiced stops > voiced stops
  - The value of glottal opening area for vowels can be set as zero
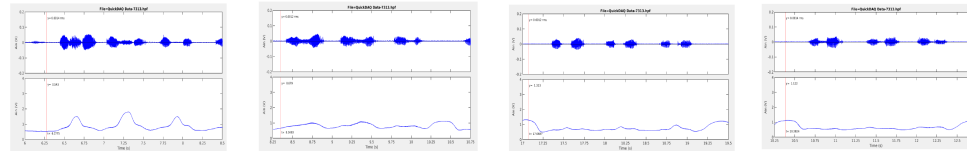

Figure 3.1: VCV pseudo-words sequence /asa-ajs-ala/, with the speech waveform above and the EPGG signals below.

Figure 3.2: VCV pseudo-words sequence /axa-aga-axa/, with the speech waveform above and the EPGG signals below.

Figure 3.3: VCV pseudo-words sequence /apa-ata-aka/, with the speech waveform above and the EPGG signals below.

Figure 3.4: VCV pseudo-words sequence /aba-ada-aga/, with the speech waveform above and the EPGG signals below.

Sample sentence (in French): Du coup l'oculiste tout fou dévissa sans scrupule le volant du véhicule.

(English translation): Suddenly the crazy oculist unscrupulously unscrewed the steering wheel of the vehicle.
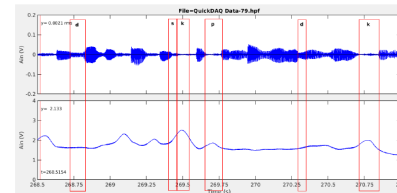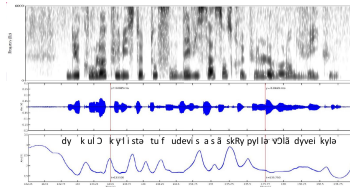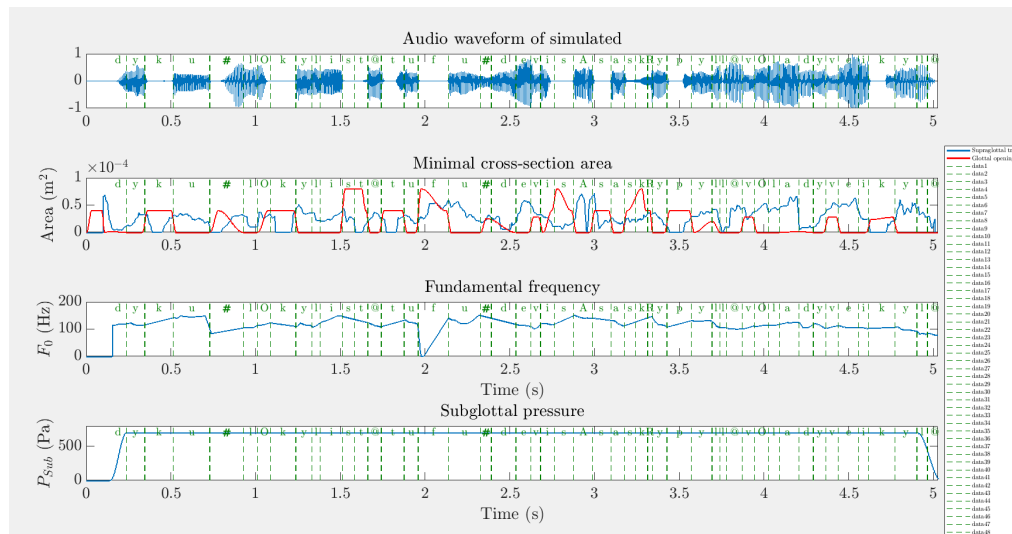

Figure 3.6: Sentence section (in French) "fou dévissa sans scrupule le volant du véhicule", with the speech waveform in the first row and the EPGG signals in the second row. The highlighted stops and fricatives are annotated by IPA.

Synthesis output

Audio waveform of simulated

Minimal cross-section area

Fundamental frequency

Subglottal pressure

EPGG system

- ElectroPhotoGlottoGraphy (EPGG) is a non-invasive technique, which mainly utilizes a light source located on the surface of the side neck to illuminate the hypopharynx and a photosensor unit located on the speaker's front neck.

- The amount of light captured is an affine function of the glottal opening area.

- By normalizing the glottal opening area to its maximal value, the glottal opening area (A_g) provided by the EPGG data can be directly related to the glottal abduction degree (D_{ab}) used in the numerical simulation (Elie and Laprie, 2017).
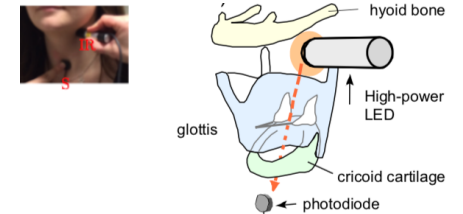
References

- Heinz, J. M. and Stevens, K. N. (1965). On the relations between lateral cineradiographs area functions and acoustic spectra of speech. In Proceedings of the 5th International Congress on Acoustics, page A44.

- Elie, B. and Laprie, Y. (2017). Acoustic impact of the gradual glottal abduction degree on the production of fricatives: A numerical study. The Journal of the Acoustical Society of America, 142(3):1303–1317.

- Laprie, Y. (2017). An articulatory model of the complete vocal tract derived from medical images. In ESSV 2017, Saarland University, Germany.

- Laprie, Y., Elie, B., Amelot, A., Vuissoz, P.-A., Maeda, S., and Demolin, D. (2019). Making the images of the vocal tract speak: Lessons and perspectives. In Embodied Speech International Workshop, Université Sorbonne, Paris, France.

Acknowledgements