ICVPB 2020
December 2-4, 2020
Grenoble, France
An Online event

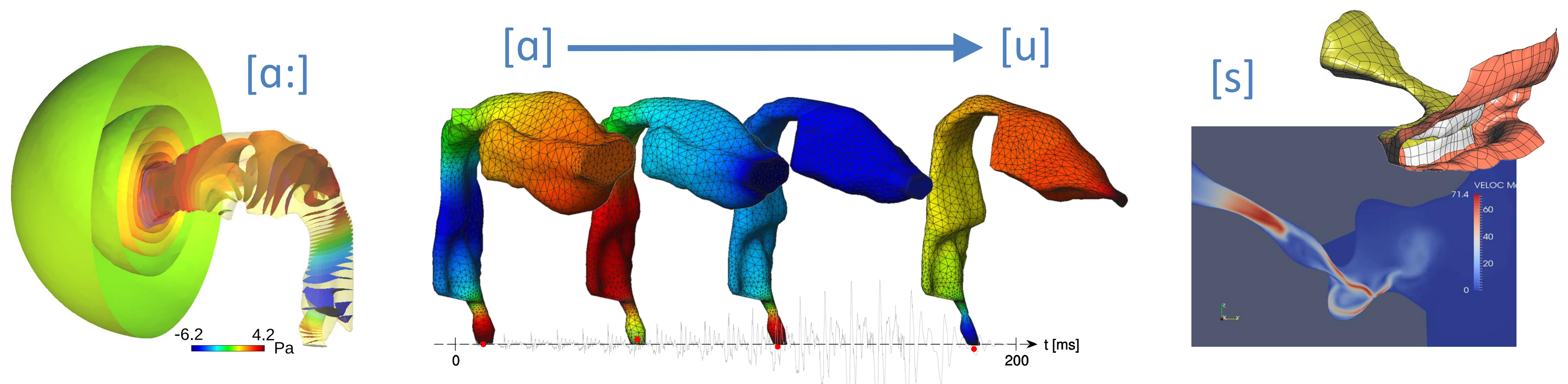
Tuning MRI-based vocal tracts to modify formants in the three-dimensional finite element production of vowels

Marc Arnela, Oriol Guasch and Arnau Pont
GTM – Grup de Recerca en Tecnologies Mèdia, La Salle – Universitat Ramon Llull, Barcelona, Catalonia, Spain
marc.arnela@salle.url.edu

laSalle Universitat Ramon Llull

1 – Introduction

The Finite Element Method (FEM) has currently been used to simulate vowels [1,2], diphthongs [3], and fricatives [4], among other sounds.



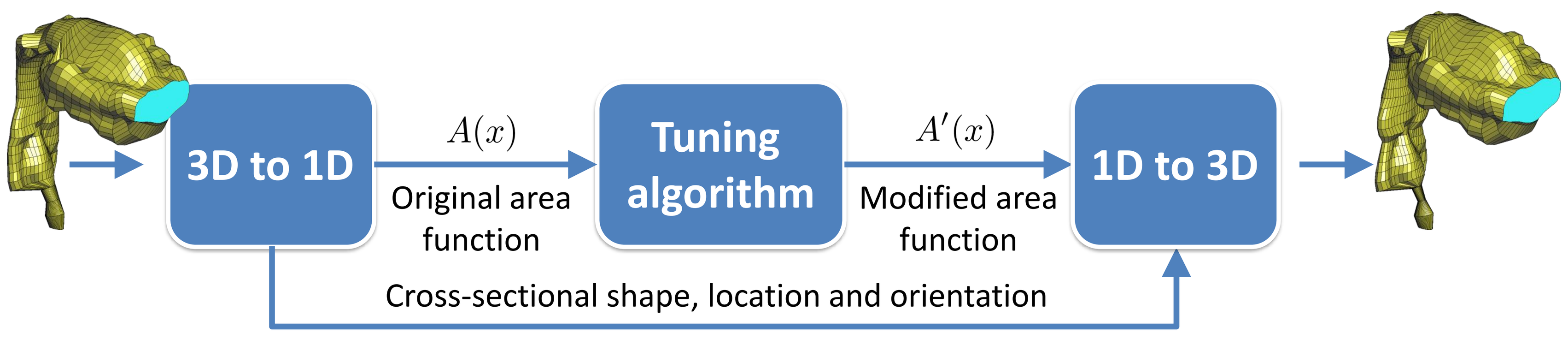
It requires accurate three-dimensional (3D) vocal tract (VT) geometries, typically obtained from Magnetic Resonance Imaging (MRI) [5]. However, one may need to slightly modify these VT geometries before running a simulation, to match, for instance, the formant frequencies obtained from natural recordings.

Goal

To modify 3D MRI-based vocal tract geometries to match target formant frequencies in the FEM simulation of vowels.

2 – Tuning methodology

Cross-sections are first extracted from the MRI-based VT [5] with an adaptive grid [2], next the area of each cross-section (VT area function) is tuned, and finally the 3D VT is reconstructed using the new areas.



Tuning algorithm [6]

The sensitivity function of the i-th formant frequency to area perturbations is computed using the FEM as

$$S_i(n) = \frac{\mathrm{KE}_i(n) - \mathrm{PE}_i(n)}{\mathrm{TE}_i} \qquad \mathrm{TE}_i = \sum_{n=1}^{N_{\text{cross}}} [\mathrm{KE}_i(n) + \mathrm{PE}_i(n)]$$

with

$$\mathrm{KE}_i(n) = \sum_a \sum_b \frac{1}{2\omega_i^2 \rho_0} \Psi_i^a \left[\int_{x_{n-1}}^{x_n} A(x) \frac{\mathrm{d}N^a}{\mathrm{d}x} \frac{\mathrm{d}N^b}{\mathrm{d}x} \,\mathrm{d}x \right] \Psi_i^b \quad \text{Kinetic Energy}$$

$$\mathrm{PE}_i(n) = \sum_a \sum_b \frac{1}{2\rho_0 c_0^2} \Psi_i^a \left[\int_{x_{n-1}}^{x_n} A(x) N^a N^b \,\mathrm{d}x \right] \Psi_i^b \quad \text{Potential Energy}$$

Ψ_i is the eigenvector with corresponding eigenfrequency ω_i resulting from the FEM resolution of the Webster's equation eigenvalue problem

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(A \frac{\mathrm{d}\bar{p}}{\mathrm{d}x} \right) + A k_0^2 \bar{p} = 0 \quad \xrightarrow{\text{FEM}} \quad -\boldsymbol{\omega}_i^2 \boldsymbol{M} \boldsymbol{\Psi}_i + \boldsymbol{K} \boldsymbol{\Psi}_i = \boldsymbol{0}$$

with

$$M_{ab} = \frac{1}{\rho_0 c_0^2} \int_0^L A(x) N^a N^b \,\mathrm{d}x \qquad K_{ab} = \frac{1}{\rho_0} \int_0^L A(x) \frac{\mathrm{d}N^a}{\mathrm{d}x} \frac{\mathrm{d}N^b}{\mathrm{d}x} \,\mathrm{d}x$$

The area function is modified following the iterative process in [7] to reach the desired target frequencies

$$A_{k+1}(n) = A_k(n) + \sum_{i=1}^{N_\omega} z_{i_k} S_{i_k}(n) \qquad z_{i_k} = \alpha \frac{\omega_i^{\text{tg}} - \omega_{i_k}}{\omega_{i_k}}$$

3 – Finite element simulations

We have used an in-house FEM code that solves the **mixed wave equation** for the acoustic pressure $p(\boldsymbol{x}, t)$ and acoustic particle velocity $\boldsymbol{u}(\boldsymbol{x}, t)$ [3]

$$\frac{1}{\rho_0 c_0^2} \partial_t p + \nabla \cdot \boldsymbol{u} = 0,$$

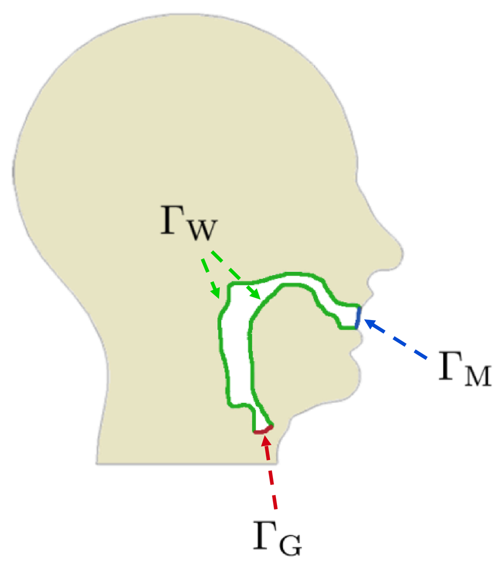
$$\rho_0 \partial_t \boldsymbol{u} + \nabla p = 0,$$

with boundary conditions

$$\boldsymbol{u} \cdot \boldsymbol{n} = g(t) \quad \text{on } \Gamma_{\mathrm{G}}$$

$$\boldsymbol{u} \cdot \boldsymbol{n} = p / Z_w \quad \text{on } \Gamma_{\mathrm{W}}$$

$$p = 0 \quad \text{on } \Gamma_{\mathrm{M}}$$

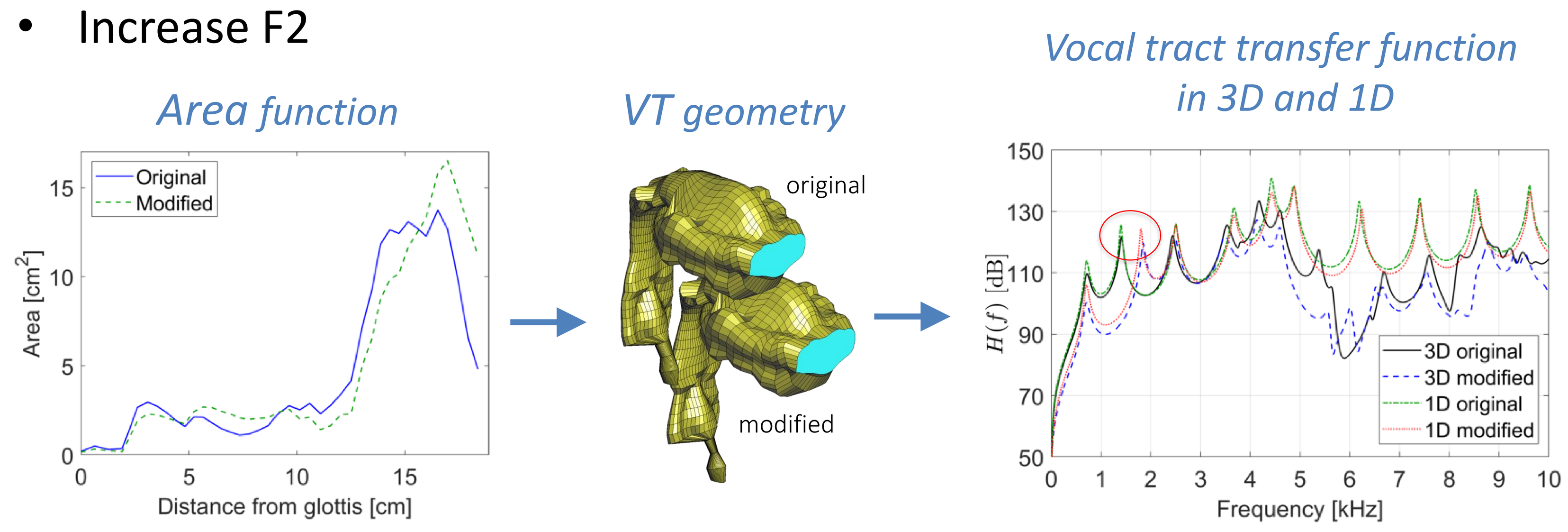


The vocal tract transfer function $H(f) = P_o(f)/Q_i(f)$ was computed for each configuration, with $P_o(f)$ being the Fourier transform of the acoustic pressure tracked at the VT exit (mouth) and $Q_i(f)$ the Fourier transform of the volume velocity introduced at the VT entrance (glottis).

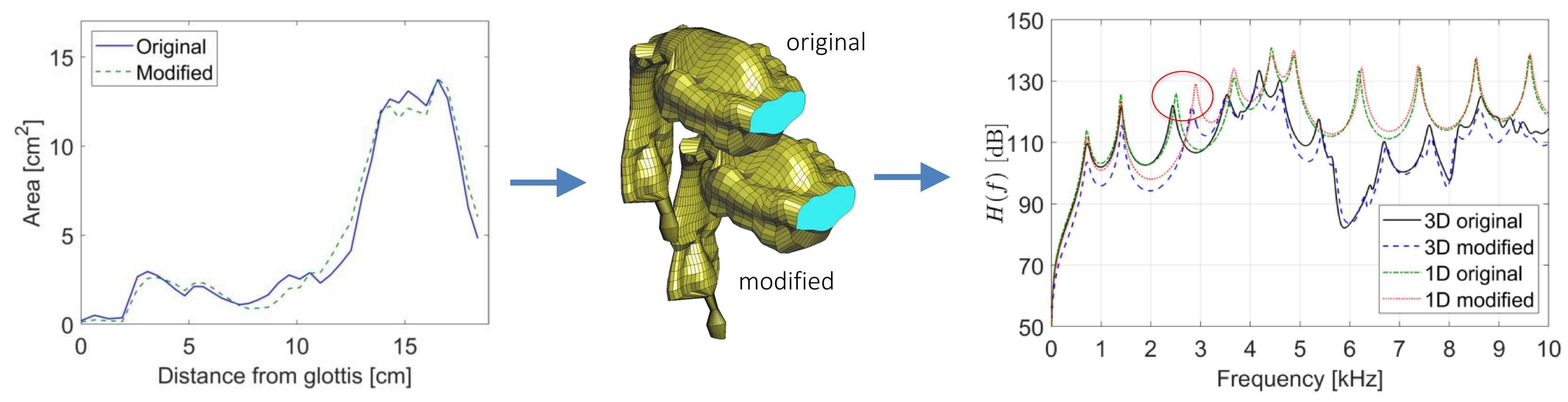
4 – Results

The vocal tract geometry of vowel [ɑ] is modified to

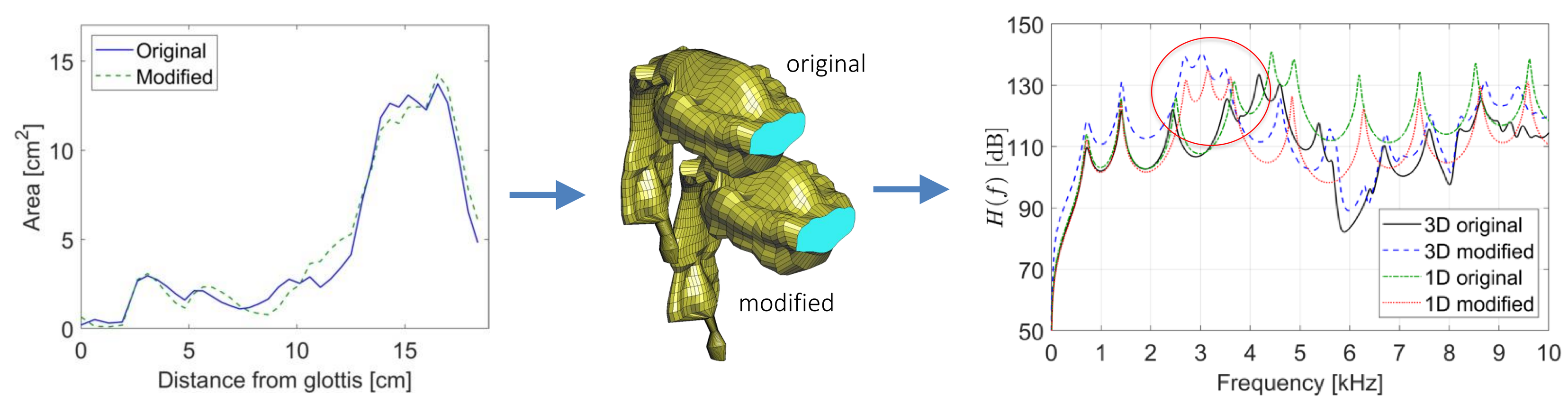
- Increase F2



- Increase F3



- Cluster F3, F4 and F5



5 – Conclusions

In this work, we have presented a methodology to tune the formants of 3D MRI-based vocal tracts. It has been applied to the FEM generation of vowel sounds. Switching between 3D and 1D allows one to modify planar mode resonances, while keeping the higher frequency content of voice dictated by the 3D VT geometry.

References and acknowledgements

[1] M. Arnela et al (2016), "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," J. Acoust. Soc. Am., 140 (3)
[2] S. Dabbaghchian, M. Arnela, O. Engwall, O. Guasch (2019), "Reconstruction of Vocal Tract Geometries from Biomechanical Simulations", Int. J. Numer. Meth. Biomed. Eng., 35 (2)
[3] M. Arnela, S. Dabbaghchian, O. Guasch, O. Engwall (2019), "MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs", IEEE/ACM TASLP, 27 (12)
[4] A. Pont, O. Guasch, M. Arnela (2020), "Finite element generation of sibilants /s/ and /z/ using random distributions of Kirchhoff's vortices". Int. J. Numer. Meth. Biomed. Eng., 36 (2)
[5] D. Aalto et al (2014),"Large scale data acquisition of simultaneous MRI and speech," Appl. Acoust., 83
[6] O. Guasch, M. Arnela, A. Pont (2020), "Resonance tuning in vocal tract acoustics from modal perturbation analysis instead of nonlinear radiation pressure", J. Sound Vib., In Press
[7] B. H. Story (2006), Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions, J. Acoustic. Soc. Am., 109 (4)

This research has been supported by project GENIOVOX TEC2016-81107-P, by URL project 2019-URL-Proj-067, and by URL project 2020-URL-Proj-056.