23rd International Congress on Sound & Vibration
Athens, Greece 10-14 July 2016
ICSV23

FINITE ELEMENT GENERATION OF VOWEL SOUNDS USING DYNAMIC COMPLEX THREE-DIMENSIONAL VOCAL TRACTS

Marc Arnela and Oriol Guasch
GTM - Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Catalonia, Spain
email: marnela@salleurl.edu

Saeed Dabbaghchian and Olov Engwall
Department of Speech, Music and Hearing, School of Computer Science and Communication,
KTH Royal Institute of Technology, Stockholm, Sweden

Three-dimensional (3D) numerical simulations of the vocal tract acoustics require very detailed vocal tract geometries in order to generate good quality vowel sounds. These geometries are typically obtained from Magnetic Resonance Imaging (MRI), from which a volumetric representation of the complex vocal tract shape is obtained. Static vowel sounds can then be generated using a finite element code, which simulates the propagation of acoustic waves through the vocal tract when a given train of glottal pulses is introduced at the glottal cross-section. A more challenging problem to solve is that of generating dynamic vowel sounds. On the one hand, the acoustic wave equation has to be solved in a computational domain with moving boundaries, which entails some numerical difficulties. On the other hand, the finite element meshes where acoustic wave propagation is computed have to move according to the dynamics of these very complex vocal tract shapes. In this work this problem is addressed. First, the acoustic wave equation in mixed form is expressed in an Arbitrary Lagrangian-Eulerian (ALE) framework to account for the vocal tract wall motion. This equation is numerically solved using a stabilized finite element approach. Second, the dynamic 3D vocal tract geometry is approximated by a finite set of cross-sections with complex shape. The time-evolution of these cross-sections is used to move the boundary nodes of the finite element meshes, while inner nodes are computed through diffusion. Some dynamic vowel sounds are presented as numerical examples.

1. Introduction

Dynamic vowel sounds such as diphthongs or hiatus have been tradicionally generated using one-dimensional (1D) approaches. To do so, the three-dimensional (3D) vocal tract geometry, typically obtained from Magnetic Resonance Imaging (MRI), is approximated by the so-called vocal tract area functions [1], which describe the changing area of the vocal tract along its midline. These area functions can be used to generate, for instance, a diphthong by interpolating the starting and ending vowel area functions. However, 1D techniques assume plane wave propagation, which holds up to about 5 kHz. Beyond this limit higher order modes can propagate [2], which can not be obviously captured by 1D methods. Three-dimensional (3D) approaches can overcome this limitation and directly work with MRI-based geometries. However, they have been mainly focused on analyzing the vocal tract acoustics of static sounds [3, 4, 5, 6, 7, 8], paying little attention to the production of diphthongs or hiatus. It is to be mentioned that some recent attempts have been done towards this direction, in [9] using the tuned 2D vocal tracts in [10] or in [11, 12] using 3D vocal tracts. However, the latter mainly focused on the numerical formulation and therefore made use of simplified straight

vocal tracts with circular cross-sections, which can be easily driven by area interpolation (similar to 1D).

In this work, dynamic vowel sounds are numerically generated by using 3D complex vocal tracts obtained from MRI. However, it is infeasible to obtain time-evolving MRI-based vocal tracts with high spatial and time resolution. Moreover, it would require image segmentation and surface reconstruction at each time step with the intervention of an expert. To surpass these limitations it is herein proposed to use static MRI-based vocal tract geometries and extract not only the vocal tract area function, but also the shape of each cross-section and its location and orientation in the midline. By using these parameters (shape, location and orientation) the 3D vocal tract can be easily reconstructed and interpolated so as to generate a dynamic vocal tract. The second problem to solve consists on simulating the propagation of acoustic waves in a moving vocal tract. To do so, the mixed wave equation for the acoustic pressure and acoustic particle velocity is expressed in an Arbitrary Lagrangian-Eulerian (ALE) frame of reference and numerically solved using the Finite Element Method (FEM). The finite element meshes are deformed in time according to the 3D dynamic vocal tract model. In particular, this vocal tract model is used to prescribe the motion of the nodes located at the vocal tract walls, while the position of the inner nodes is obtained from the solution to the Laplacian equation, which smoothly translates the movement of the boundary nodes to the inner ones through diffusion.

This work is organized as follows. Section 2 describes the methodology followed to generate dynamic vowel sounds using 3D-FEM and 3D MRI-based vocal tracts. Some numerical results of diphthong sounds are next presented in Section 3. Section 4 closes the paper with the conclusions.

2. Methodology

2.1 The dynamic 3D vocal tract model

The MRI-based vocal tract geometries for vowel sounds in [13] were used as a starting point. First, they were adapted for our purposes by removing the face and neck, the lips and the subglottal tube. Then, a set of cross-sections were extracted through the vocal tract midline for each vowel vocal tract geometry as in [14]. The next usual step would have been that of computing the area of each cross-section so as to obtain the so-called vocal tract area functions (see e.g., [1]), typical from 1D approaches. However, since our aim is to deal with 3D complex vocal tract geometries, the shape of each cross-section was preserved together with its location and orientation through the vocal tract midline. The vocal tract geometry can then be reconstructed by connecting each cross-section so as to form a quad-faced surface mesh. An example of this procedure is illustrated in Fig. 1, where the MRI-based vocal tract geometry of vowel [ɑ] is discretized in 40 cross-sections and then reconstructed.

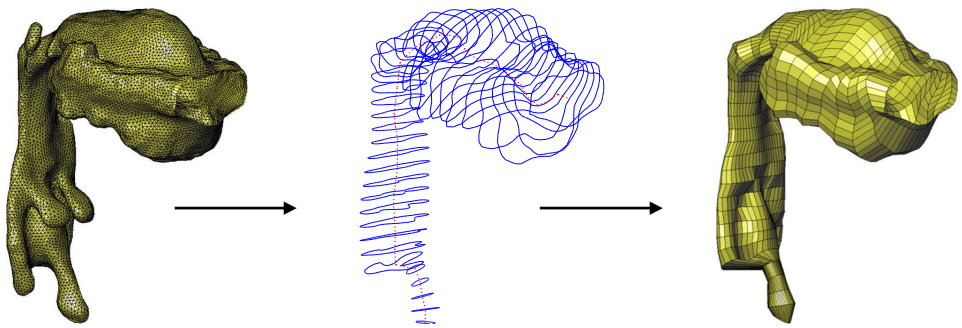

Figure 1: (Left) MRI-based vocal tract geometry for vowel [ɑ], (mid) corresponding 2D cross-sections (solid blue line) represented through the 3D vocal tract midline (dotted red line), and (right) reconstructed vocal tract geometry.

Once the cross-section database with the different vowel vocal tract geometries is constructed, a dynamic vocal tract geometry can be simply generated by interpolation. To that purpose, all cross-sections are interpolated in time from the initial vowel sound to the target one, as well as their location and orientation in the midline. Note that with this methodology we have reduced a 3D interpolation problem, which would have required working with 3D complex geometries, to a 2D interpolation problem for the cross-sections and a 1D interpolation problem for the location and orientation of each cross-section. Figure 2 shows an example of the dynamic vocal tract model when moving from vowel [ɑ] to vowel [i]. It can be appreciated how each cross-section of the vocal tract deforms to reach the desired articulation, and moves along the midline changing its orientation to correctly represent the vocal tract geometry. The equivalent surface meshes could have also been generated at any time instant, as exemplified in Fig. 1 for vowel [ɑ]. However, this option is only required at the first time step so as to generate an initial surface mesh. As it will be explained in the Section 2.3, this initial geometry will be meshed so as to obtain a volumetric finite element mesh to simulate acoustic wave propagation. The boundaries of this volumetric mesh will move according to the variations of each cross-section.

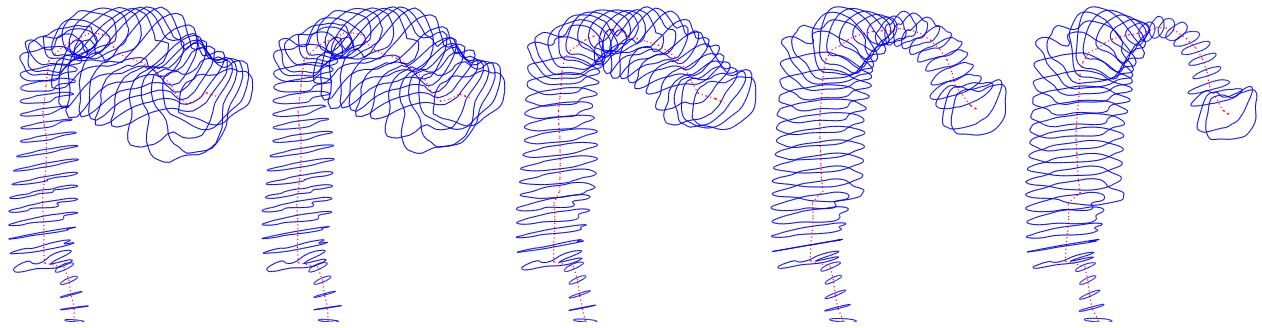

Figure 2: Dynamic vocal tract model moving from vowel [ɑ] to vowel [i].

2.2 The acoustic wave equation for dynamic vowel sounds

The vocal tract acoustics for static vowel sounds can be modeled using the mixed wave equation for the acoustic pressure $p(\boldsymbol{x},t)$ and acoustic particle velocity $\boldsymbol{u}(\boldsymbol{x},t)$,

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

$$\rho_0\partial_t\boldsymbol{u} + \nabla p = 0, \tag{1b}$$

with ∂_t denoting the partial time derivative and ρ_0 and c_0 respectively standing for the air density and the speed of sound. However, when facing the modelling of dynamic vowel sounds such as diphthongs or hiatus one has to deal with moving vocal tracts. In such a situation it becomes necessary to express the wave equation (1) in an Arbitrary Lagrangian-Eulerian (ALE) frame of reference. One option is to resort to a quasi-Eulerian ALE framework [15, 16], where the spatial derivatives are kept in an Eulerian frame and the time derivatives are translated into a referential frame moving with the domain. Consider that the domain moves with a velocity $\boldsymbol{u}_{\text{dom}}$. The ALE mixed wave equation can be obtained by replacing $\partial_t f \leftarrow \partial_t f - \boldsymbol{u}_{\text{dom}}\cdot\nabla f$ in Eq. (1), which results in

$$\frac{1}{\rho_0 c_0^2}\partial_t p - \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}}\cdot\nabla p + \nabla\cdot\boldsymbol{u} = 0, \tag{2a}$$

$$\rho_0\partial_t\boldsymbol{u} - \rho_0\boldsymbol{u}_{\text{dom}}\cdot\nabla\boldsymbol{u} + \nabla p = 0. \tag{2b}$$

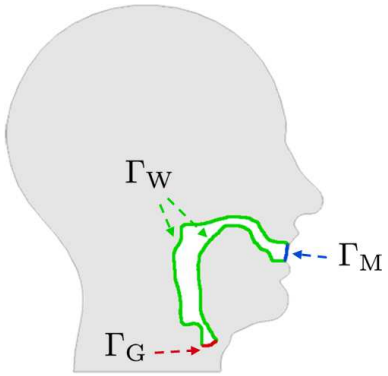

Figure 3: A sketch of the computation domain Ω showing the glottal boundary Γ_{G}, the vocal tract walls Γ_{M} and the mouth exit Γ_{M}.

In order to generate a vowel sound one has to supplement Eq. (2) with appropriate boundary and initial conditions. Let us consider a computational domain Ω representing the vocal tract geometry with boundaries Γ_{W}, Γ_{M} and Γ_{G} respectively standing for the vocal tract walls, the mouth exit and the glottal cross-section where the vocal folds are located (see Figure 3). We impose

$$\boldsymbol{u}\cdot\boldsymbol{n} = g(t) \qquad \text{on } \Gamma_{\mathrm{G}},\ t>0, \tag{3a}$$
$$\boldsymbol{u}\cdot\boldsymbol{n} = p/Z_w \qquad \text{on } \Gamma_{\mathrm{W}},\ t>0, \tag{3b}$$
$$p = 0 \qquad \text{on } \Gamma_{\mathrm{M}},\ t>0, \tag{3c}$$
$$p = 0,\ \boldsymbol{u} = 0 \qquad \text{in } \Omega,\ t=0, \tag{3d}$$

where $g(t)$ is the glottal source inflow and Z_w the wall impedance used to introduce wall losses. For simplicity, a zero pressure release condition is used in Eq. (3c) to consider an open-end boundary condition. However, radiation losses could be easily introduced by extending the computational domain outside the vocal tract geometry and considering free-field sound propagation (see e.g., [5, 7]), although at the cost of increasing the computation time.

2.3 Time domain finite element simulations

Acoustic wave propagation through the 3D dynamic vocal tracts was simulated by using the Finite Element Method to solve the ALE mixed wave equation (2) with the boundary and initial conditions in (3). In particular, an algebraic subgrid scale strategy was followed [12], which allows us to prevent numerical instabilities when using the same interpolation for the acoustic pressure and particle velocity. A train of glottal pulses of the Rosenberg type [17] was generated and introduced at the vocal tract entrance as $g(t)$ in Eq. (3a). This train of pulses was enhanced by considering a pitch curve (variation of fundamental frequency), some jitter and shimmer and a fade in/out to emulate the onset/offset of the vocal folds. Wall losses were considered in Eq. (3b) by imposing a wall impedance of $Z_w = 83666\ \mathrm{kg/m^2s}$ [3]. A speed of sound of $c_0 = 350$ m/s and an air density of $\rho_0 = 1.14\ \mathrm{kg/m^3}$ were used. A sampling frequency of $f_s \equiv 1/\Delta t = 250$ kHz was selected to simulate time events of $T = 200$ ms, with Δt denoting the time step size.

In what concerns the motion of the finite element meshes, we depart from an intermediate vocal tract geometry corresponding to the average between the initial and final vowel sound to generate, instead of starting from the initial vowel vocal tract geometry. This helped us to minimize element distortion produced when deforming the finite element mesh to generate a dynamic vowel sound, and thus to avoid remeshing strategies. This initial geometry was meshed using tetrahedral elements of size $h \sim 0.003$ m. Its boundary nodes were moved to reach the first vowel sound and then to the second vowel sound so as to generate a diphthong or a hiatus. The trajectories of these boundary nodes were obtained by using the dynamic vocal tract model described in Section 2.1. The position

of the inner nodes were then computed through diffusion. In particular, the Finite Element Method was employed to solve the Laplacian equation for the node displacements $\boldsymbol{w}(\boldsymbol{x}, t)$,

$$\nabla^2 \boldsymbol{w}^{n+1} = 0 \qquad \text{in } \Omega,\ t = t^{n+1}, \tag{4a}$$

with boundary conditions

$$\boldsymbol{w}^{n+1} = \boldsymbol{x}^{n+1}_{\text{walls}} - \boldsymbol{x}^{n}_{\text{walls}} \qquad \text{on } \Gamma_{\text{W}},\ t = t^{n+1}, \tag{4b}$$

$$\boldsymbol{w}^{n+1} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{G}},\ t = t^{n+1}, \tag{4c}$$

$$\boldsymbol{w}^{n+1} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{M}},\ t = t^{n+1}, \tag{4d}$$

where $\boldsymbol{x}^{n+1}_{\text{walls}}$ and $\boldsymbol{x}^{n}_{\text{walls}}$ stand for the positions of the nodes located in the vocal tract walls at time instants $n+1$ and n, respectively. The node positions at the $n+1$ time instant can be updated as

$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^{n} + \boldsymbol{w}^{n+1}. \tag{5}$$

The mesh velocity $\boldsymbol{u}_{\text{dom}}$ appearing in the ALE mixed wave equation (2) is computed at a point $\boldsymbol{x}_i$ as

$$\boldsymbol{u}^{n+1}_{\text{dom}}(\boldsymbol{x}_i) = \frac{\boldsymbol{x}^{n+1}_i - \boldsymbol{x}^{n}_i}{\Delta t}. \tag{6}$$

3. Results

The production of diphthong [ɑi] has been simulated as an example. As detailed in Section 2.3, we started from an initial finite element mesh corresponding to an intermediate position between vowel [ɑ] and [i]. This initial mesh was then deformed to reach the articulation of vowel [ɑ]. This procedure corresponded to a 150 ms simulation and only involved the resolution of the Laplacian equation (4). Once the articulation of the initial vowel was reached, the ALE mixed wave equation (2) with boundary conditions (3) was numerically solved to simulate acoustic wave propagation through the

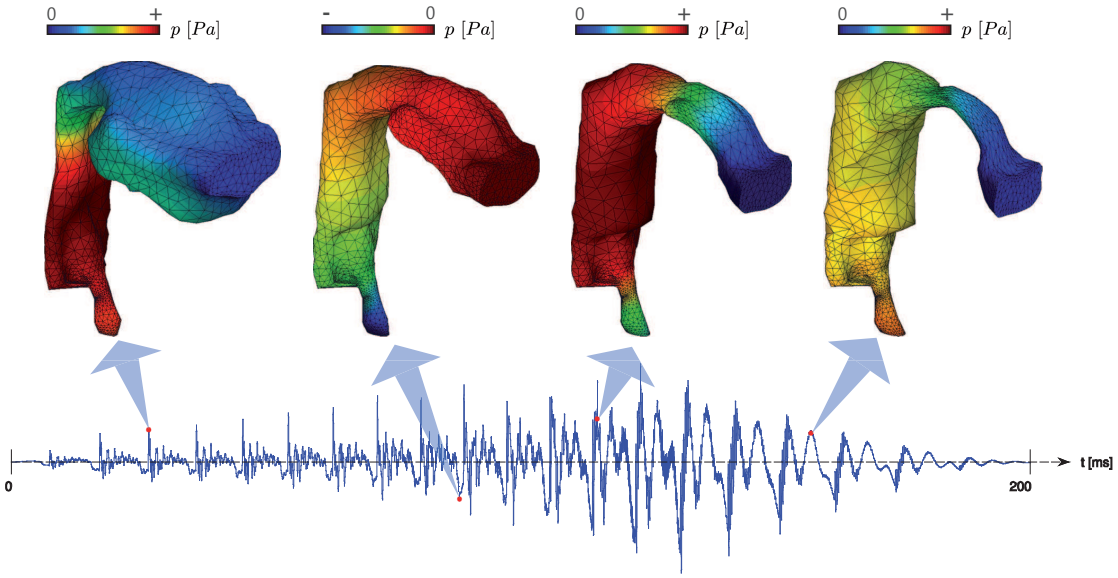

Figure 4: Snapshots showing the acoustic pressure distribution of the vocal tract walls during the numerical simulation of diphthong [ɑi] (top) with the time evolution of the acoustic pressure captured at the vocal tract exit (bottom). Values were taken at t=(27, 88, 115, 157) ms, respectively corresponding to the articulation of vowel [ɑ], two intermediate positions between vowels [ɑ] and [i], and to the articulation of vowel [i].

vocal tract, together with the Laplacian equation (4) used to move the finite element meshes. The articulations of vowel [ɑ] and vowel [i] were respectively sustained during 15 ms and 35 ms, while the transition time between vowels lasted 150 ms. This gives a total time of 200 ms for the generation of the diphthong [ɑi].

Figure 4 presents a set of snapshots showing the acoustic pressure distribution of the vocal tract walls at different time instants. The first one corresponds to the articulation of vowel [ɑ] (t=27 ms), the second and third ones to intermediate positions between [ɑ] and [i] (t=88 ms and t=115 ms), and the last one to the articulation of vowel [i] (t=157 ms). The time evolution of the acoustic pressure captured at a node close to the mouth exit is also represented in the figure to better illustrate theses time instants. It can be observed how the vocal tract geometry smoothly moves from the articulation of vowel [ɑ] to that of vowel [i], producing a smooth variation in time of the acoustic pressure captured at the vocal tract exit. This signal indeed corresponds to the produced diphthong sound, which can be transformed to an audio file to listen to it. The spectrogram of this audio signal is represented in Figure 5. As usually done in the literature, a pre-emphasis filter was applied to better visualize the higher frequencies. We can observe in the figure how the formants smoothly transition from those of vowel [ɑ] to those of vowel [i].

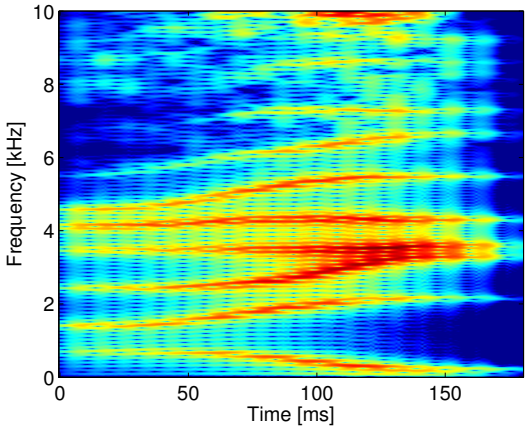

Figure 5: Spectrogram of the generated diphthong [ɑi].

4. Conclusions

In this work a methodology to generate dynamic vowel sounds using 3D MRI-based vocal tract geometries has been presented. A 3D dynamic vocal tract model has been constructed for this purpose, which can generate the articulation of diphthongs and hiatus by interpolating the shape, location and orientation of a set of cross-sections extracted from static MRI-based geometries of vowels. The Finite Element Method has been used to simulate acoustic wave propagation through these dynamic vocal tracts. To do so, the mixed wave equation for the acoustic pressure and particle velocity has been set in an ALE framework and solved following a subgrid scale strategy. The boundary nodes of the finite element meshes corresponding to the vocal tract walls have been moved according the the constructed 3D dynamic vocal tract model, while inner node positions have been obtained by solving the Laplacian equation. As a numerical example the diphthong [ɑi] has been simulated, showing a smooth transition of the formants during its production.

5. Acknowledgements

This research has been supported by EU-FET grant EUNISON 308874.

REFERENCES

1. Story, B. H., Titze, I. R. and Hoffman, E. A. Vocal tract area functions from magnetic resonance imaging, *J. Acoust. Soc. Am.*, **100**(1), 537–554, (1996).

2. Blandin, R., Arnela, M., Laboissière, R., Pelorson, X., Guasch, O., Van Hirtum, A. and Labal, X. Effects of higher order propagation modes in vocal tract like geometries, *J. Acoust. Soc. Am.*, **137**(2), 832–843, (2015).

3. Švancara, P. and Horáček, J. Numerical modelling of effect of tonsillectomy on production of czech vowels, *Acta Acust. united with Acustica*, **92**(5), 681–688, (2006).

4. Vampola, T., Horáček, J. and Švec, J. G. FE modeling of human vocal tract acoustics. Part I: Production of czech vowels, *Acta Acust. united with Acustica*, **94**(5), 433–447, (2008).

5. Takemoto, H., Mokhtari, P. and Kitamura, T. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method, *J. Acoust. Soc. Am.*, **128**(6), 3724–3738, (2010).

6. Arnela, M. and Guasch, O. Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method, *J. Acoust. Soc. Am.*, **133**(6), 4197–4209, (2013).

7. Arnela, M., Guasch, O. and Alías, F. Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations, *J. Acoust. Soc. Am.*, **134**(4), 2946–2954, (2013).

8. Arnela, M., Blandin, R., Dabbaghchian, S., Guasch, O., Alías, F., Pelorson, X., Van Hirtum, A. and Engwall, O. Influence of lips on the production of vowels based on finite element simulations and experiments, *J. Acoust. Soc. Am.*, **139**(5), 2852–2859, (2016).

9. Arnela, M., Guasch, O., Codina, R. and Espinoza, H. Finite element computation of diphthong sounds using tuned two-dimensional vocal tracts, *Proc. of 7th Forum Acousticum*, Kraków, Poland, Setember, (2014).

10. Arnela, M. and Guasch, O. Two-dimensional vocal tracts with three-dimensional behaviour in the numerical production of vowels, *J. Acoust. Soc. Am.*, **135**(1), 369–379, (2014).

11. Guasch, O., Arnela, M., Codina, R. and Espinoza, H. Stabilized finite element formulation for the mixed convected wave equation in domains with driven flexible boundaries, *Noise and Vibration: Emerging Technologies (NOVEM2015)*, Dubrovnik, Croatia, April, (2015).

12. Guasch, O., Arnela, M., Codina, R. and Espinoza, H. A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production, *Acta Acust. united with Acustica*, **102**(1), 94–106, (2016).

13. Aalto, D., et al. Large scale data acquisition of simultaneous MRI and speech, *Appl. Acoust.*, **83**, 64–75, (2014).

14. Dabbaghchian, S., Arnela, M. and Engwall, O. Simplification of vocal tract shapes with different levels of detail, *Proc. of 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, Scotland, UK, August, (2015).

15. Hughes, T. J. R., Liu, W. K. and Zimmermann, T. K. Lagrangian-eulerian finite-element formulation for compressible viscous flows, *Comput. Methods Appl. Mech. Engrg.*, **29**(3), 329–349, (1981).

16. Huerta, A. and Liu, W. K. Viscous flow with large free surface motion, *Comput. Methods Appl. Mech. Engrg.*, **69**(3), 277–324, (1988).

17. Rosenberg, A. E. Effect of glottal pulse shape on the quality of natural vowels, *J. Acoust. Soc. Am.*, **49**(2), 583–590, (1971).