

A semi-polar grid strategy for the three-dimensional finite element simulation of vowel-vowel sequences

Marc Arnela¹, Saeed Dabbaghchian², Oriol Guasch¹, Olov Engwall²

¹GTM - Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Catalonia, Spain

²Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

marnela@salleurl.edu, saeedd@kth.se, oguasch@salleurl.edu, engwall@kth.se

Abstract

Three-dimensional computational acoustic models need very detailed 3D vocal tract geometries to generate high quality sounds. Static geometries can be obtained from Magnetic Resonance Imaging (MRI), but it is not currently possible to capture dynamic MRI-based geometries with sufficient spatial and time resolution. One possible solution consists in interpolating between static geometries, but this is a complex task. We instead propose herein to use a semi-polar grid to extract 2D cross-sections from the static 3D geometries, and then interpolate them to obtain the vocal tract dynamics. Other approaches such as the adaptive grid have also been explored. In this method, cross-sections are defined perpendicular to the vocal tract midline, as typically done in 1D to obtain the vocal tract area functions. However, intersections between adjacent cross-sections may occur during the interpolation process, especially when the vocal tract midline quickly changes its orientation. In contrast, the semi-polar grid prevents these intersections because the plane orientations are fixed over time. Finite element simulations of static vowels are first conducted, showing that 3D acoustic wave propagation is not significantly altered when the semi-polar grid is used instead of the adaptive grid. The vowel-vowel sequence [ai] is finally simulated to demonstrate the method.

Index Terms: voice production, vocal tract acoustics, Finite Element Method, semi-polar grid, dynamic vocal tract, vowel-vowel sequences, speech synthesis.

1. Introduction

Three-dimensional (3D) computational acoustic models allow for the simulation of 3D acoustic waves propagating through the intricate vocal tract and radiating to free-field space. The production of vowel sounds has been largely studied using either the Finite Element Method (FEM) in the time domain [1, 2, 3, 4] or in the frequency domain [5, 6, 7], finite differences [8, 9, 10], multimodal approaches [11, 12], or digital wave guide models [13, 14]. All these models require very detailed 3D vocal tract geometries to achieve high quality vowel sounds. For static vowels these geometries can be obtained, for instance, from Magnetic Resonance Imaging (MRI) [15]. However, a problem arises when facing the simulation of dynamic vowel sequences. In such a situation dynamic 3D vocal tract geometries cannot be obtained from MRI, since that would require the real-time image acquisition of several slices with a very high spatial resolution, which is not yet feasible.

To circumvent that problem, one can resort to biomechanical models that could reconstruct the vocal tract airway [16],

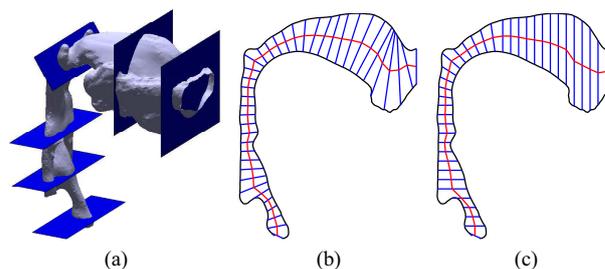


Figure 1: (a) MRI-based vocal tract geometry of vowel [a] with some gridplanes, and intersection of the gridplanes (blue) with the vocal tract boundary (black) using (b) the adaptive grid and (c) the semi-polar grid. The midline is also shown (red).

or to use interpolation between already known static geometries to define the vocal tract movement. The latter was the option adopted in [17], but instead of directly interpolating the 3D MRI-based vocal tracts, it was proposed to interpolate 2D cross-sections extracted from the static vocal tract geometries. These cross-sections were obtained from the intersection of a set of gridplanes with the 3D vocal tract walls (see Figure 1). In particular, the adaptive grid (AG) strategy presented in [18, 19] was adopted, which extracts cross-sections perpendicular to the vocal tract midline (see Fig. 1b). This allows one to also compute the vocal tract area functions [20] typically used in 1D synthesizers [21, 22], since these planes coincide with the planar wavefronts. The interpolation problem in AG was then greatly simplified because only a few parameters describing the vocal tract geometry were needed (2D cross-sectional shape, location, and orientation in the vocal tract midline). However, that approach has some limitations. Some adjacent cross-sections may intersect during the interpolation process which requires manual intervention to fix them. This may occur, for instance, when the vocal tract midline abruptly changes its curvature. In a straightened vocal tract this would never happen because the cross-section orientations are the same for all time instants [23, 24].

An approach in which the cross-section orientation was fixed would be thus appealing for dynamic 3D acoustic simulations. A simple option could consist in using a semi-polar grid (SPG) (Figure 1c) instead of the AG (Figure 1b). The former is made of horizontal planes in the pharynx, polar planes in the velar region, and vertical planes in the oral cavity that do not change their orientation over time. However, the gridplanes do not coincide with the planar wavefronts, so that 1D area functions computed using this grid would modify the sound quality

generated by 1D synthesizers. This shall not be a restriction in 3D because the wavefronts are no longer linked to the cross-sections. It is the aim of this work to show that the SPG can perform well in these situations.

This paper is organized as follows. Section 2 presents the methodology used to generate vowel-vowel sequences using the SPG. FEM in the time-domain is used to simulate 3D acoustic wave propagation within the dynamic vocal tracts. Section 3 shows the obtained results. First, the vocal tract acoustic response of vowels [a], [i] and [u] is analyzed for both the SPG and the AG. Second, the vowel-vowel sequence [ai] is generated following the SPG strategy. Conclusions close this work in Section 4.

2. Methodology

2.1. Vocal tract geometries

We have used the detailed MRI-based vocal tract geometries in [15], but employed a vocal tract geometry simplification method to resample the complex geometries [18, 19]. The main idea is to represent a 3D geometry with a set of 2D planar cross-sections, the shape of the geometry being preserved in each plane. However, it is to be noted that the reconstructed geometry between planes can be slightly modified, since the method uses linear interpolation between two adjacent cross-sections.

First, side branches including sinus piriformis and vallecula, and also the lips, were removed from the original geometries as shown in Figure 1a (see e.g. [8] and [25] for their influence on the generated sound). Then a set of gridplanes were defined to cut the vocal tract geometry. Figure 1a depicts some of these planes. Each plane is identified by its position and orientation, and a set of planes define a grid system, such as the SPG [18, 26], which consists of three sections. The first section starts from the glottis and ends at the oropharynx just above the vallecula; all planes are defined horizontally in this section. In the second section, which approximately covers the velar region, the orientation of the planes changes in the interval $[0, \pi/2]$ with equal steps. The third section, which ends at the mouth opening, contains vertical planes. The number of planes in each section can be adjusted. Figure 1c depicts such a grid (blue color) and the vocal tract outline (black color) in the midsagittal plane. The red curve in this figure presents the vocal tract midline.

It can be observed in Figure 1c that with the SPG the gridplanes are not perpendicular to the vocal tract midline, as one would require for extracting 1D vocal tract area functions [20]. The latter could be achieved using an AG. In such a grid, the SPG is adapted for each geometry so that all gridplanes are perpendicular to the midline. To do so, an initial midline is first calculated using the SPG and then smoothed and resampled at even intervals. The tangent of the smoothed midline at given sampling points defines the normal vectors of the gridplanes. Using these gridplanes, a final midline is calculated by intersecting the geometry with the AG. Figure 1b illustrates an example of the vocal tract midsagittal boundary (black color), orientation of the gridplanes (blue color), and midline (red color) obtained by using the AG.

Unfortunately, collision between adjacent cross-sections may occur in the AG when the number of cross-sections increases. As Figure 1b shows, the cross-sections are very close to each other at the alveolar ridge. Moreover, interpolation between static configurations of the vocal tract becomes complex when they are resampled with the AG. As an alternative, the

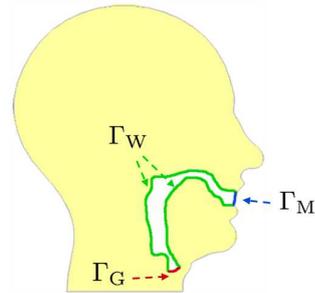


Figure 2: Sketch showing the boundaries of the computational domain Ω . Γ_G stands for the glottis, Γ_W for the vocal tract walls, and Γ_M for the mouth aperture.

SPG helps one to avoid the intersection between adjacent cross-sections, since the orientation of the gridplanes is fixed. The SPG also simplifies the interpolation between static configurations.

2.2. Finite element simulations

2.2.1. The acoustic wave equation

Static vowel sounds can be generated by solving the mixed wave equation for the acoustic pressure $p(\mathbf{x}, t)$ and acoustic particle velocity $\mathbf{u}(\mathbf{x}, t)$

$$\frac{1}{\rho_0 c_0^2} \partial_t p + \nabla \cdot \mathbf{u} = 0, \quad (1a)$$

$$\rho_0 \partial_t \mathbf{u} + \nabla p = 0, \quad (1b)$$

with ρ_0 standing for the air density and c_0 for the speed of sound. In the case of a vowel-vowel sequence, one has to express Eq. (1) in an ALE (Arbitrary Lagrangian-Eulerian) frame of reference to account for the vocal tract movement. This can be achieved by replacing $\partial_t f \leftarrow \partial_t f - \mathbf{u}_{\text{dom}} \cdot \nabla f$ in Eq. (1), with \mathbf{u}_{dom} standing for the velocity of the domain and f for an arbitrary function. This yields [23, 24]

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

$$\rho_0 \partial_t \mathbf{u} - \rho_0 \mathbf{u}_{\text{dom}} \cdot \nabla \mathbf{u} + \nabla p = 0. \quad (2b)$$

Equations (1) and (2) have to be supplemented with proper boundary and initial conditions for vowel sounds. Consider the computational domain Ω with Γ_G being the glottis, Γ_W the vocal tract walls, and Γ_M the boundary that closes the mouth aperture (see Figure 2). We introduce an acoustic particle velocity $u_g(t)$ at the vocal tract entrance Γ_G , assume wall losses with a constant frequency impedance $Z_w = 83666 \text{ kg/m}^2\text{s}$ [1], and suppose an open-end boundary condition ($p = 0$) at Γ_M . The latter does not introduce radiation losses, but they could be easily considered by allowing sound waves to radiate outside the vocal tract towards infinity (see e.g., [1, 3, 8]). The above boundary conditions read

$$\mathbf{u} \cdot \mathbf{n} = u_g(t) \quad \text{on } \Gamma_G, t > 0, \quad (3a)$$

$$\mathbf{u} \cdot \mathbf{n} = p/Z_w \quad \text{on } \Gamma_W, t > 0, \quad (3b)$$

$$p = 0 \quad \text{on } \Gamma_M, t > 0, \quad (3c)$$

$$p = 0, \mathbf{u} = 0 \quad \text{in } \Omega, t = 0. \quad (3d)$$

The Finite Element Method (FEM) has been used to solve Eq. (1) for static vowel sounds and Eq. (2) for dynamic vowel sounds with boundary and initial conditions (3). The algebraic subgrid scale strategy in [24] was implemented to use the same interpolation for the acoustic pressure and acoustic particle velocity, preventing the appearance of numerical instabilities.

2.2.2. Dynamic finite element meshes

In the particular case of vowel-vowel sequences one has to solve an additional problem to move the finite element meshes according to the vocal tract deformation. In this case the location of the nodes at the vocal tract walls $\mathbf{x}_{\text{walls}}(t)$ is known, as they are driven by the vocal tract wall movement, but the inner node positions have to be relocated accordingly. A standard option consists in solving the Laplacian equation for the node displacements \mathbf{w} , which smoothly translates the movement of the vocal tract wall nodes to the inner nodes through diffusion.

Suppose that the simulation time interval $t = [0, T]$ is discretized with a constant time step $\Delta t = t^n - t^{n-1}$, with t^n being the time instant at the n th iteration. The additional problem to solve reads

$$\nabla^2 \mathbf{w}^{n+1} = 0 \quad \text{in } \Omega, t = t^{n+1}, \quad (4a)$$

with boundary conditions

$$\mathbf{w}^{n+1} = \mathbf{x}_{\text{walls}}^{n+1} - \mathbf{x}_{\text{walls}}^n \quad \text{on } \Gamma_W, t = t^{n+1}, \quad (4b)$$

$$\mathbf{w}^{n+1} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_G, t = t^{n+1}, \quad (4c)$$

$$\mathbf{w}^{n+1} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_M, t = t^{n+1}. \quad (4d)$$

Note that Eq. (4b) stands for the vocal tract wall movement, while the displacement in the normal direction is zero in Eq. (4c) and Eq. (4d) to avoid artificial lengthening of the vocal tract.

2.2.3. Simulation details

A static vowel sound can be produced with a train of glottal pulses for $u_g(t)$ in Eq. (3a). However, a broadband signal was first used instead to examine the vocal tract acoustic response for different geometry configurations. The following Gaussian pulse was introduced at Γ_G ,

$$u_g(t) = e^{-[(t-T_{gp})/0.29T_{gp}]^2} [\text{m/s}], \quad (5)$$

with $T_{gp} = 0.646/f_c$ and $f_c = 10$ kHz. This pulse was low-pass filtered at 10 kHz to avoid the appearance of numerical errors above the maximum frequency of analysis, $f_{max} = 10$ kHz. An FEM simulation was then performed to solve Eq. (1) with boundary and initial conditions (3). A speed of sound of $c_0 = 350$ m/s was used, and the sampling frequency was set to $f_s = 1/\Delta t = 160$ kHz. A 50 ms simulation was performed capturing the acoustic pressure evolution $p_o(t)$ at a node close to the mouth. The following vocal tract transfer function was then computed

$$H(f) = \frac{P_o(f)}{Q_g(f)}, \quad (6)$$

with $P_o(f)$ standing for the Fourier transform of $p_o(t)$ and $Q_g(f)$ for the Fourier transform of the volume velocity $Q_g(t)$ introduced at the glottis. Note that the Gaussian pulse in Eq. (5) corresponds to a particle velocity, so one has to multiply it by the constant cross-section at the glottis, A_g , to obtain a volume velocity, i.e. $Q_g(t) = u_g(t)A_g$.

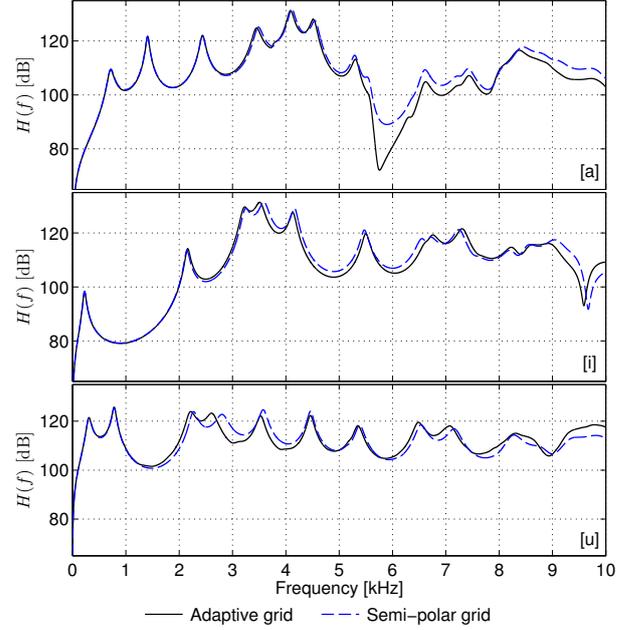


Figure 3: Vocal tract transfer functions $H(f)$ for vowels [a], [i] and [u] using the vocal tract geometries resampled with the adaptive grid and the semi-polar grid.

In the case of the vowel-vowel sequence [ai] a train of glottal pulses was used for $u_g(t)$ instead of (5) to analyze the generated sound. A Rosenberg model was selected [27]. Some modifications were introduced to enhance the naturalness of the produced sound, consisting of a pitch curve, a fade in/out, and some jitter and shimmer. Acoustic wave propagation was simulated by solving Eq. (2) with boundary and initial conditions (3), while finite element meshes were moved according to Eq. (4). The coordinates of the boundary nodes $\mathbf{x}_{\text{walls}}(t)$ were determined by linearly interpolating the cross-sections extracted from each static vocal tract geometry with the SPG method (see Section 2.1). In order to minimize the distortion of the elements during the numerical simulation, we started from an initial finite element mesh corresponding to an intermediate vocal tract geometry of the vowel-vowel sequence [ai]. Linear tetrahedral elements of size ~ 0.005 m were used. This initial mesh was moved to the starting vowel [a] and then to [i] to generate [ai]. This strategy allowed us to avoid remeshing approaches for large element distortion, which are very time consuming and not easy to implement. The sampling frequency was increased to $f_s = 250$ kHz to make the simulations more robust to small and distorted elements. Finite element meshes were moved at a rate of $f_{mesh} = 1$ kHz, high enough to capture the vocal tract movement.

3. Results

3.1. Vowels [a], [i] and [u]

The vocal tract transfer functions $H(f)$ defined in Eq. (6) were computed for vowels [a], [i] and [u]. Two different discretizations of the original MRI-based vocal tract geometries were considered. In the first one cross-sections were obtained using the AG with planes perpendicular to the vocal tract midline, whereas for the second one cross-sections were extracted using the SPG (see Figure 1). Figure 3 shows the obtained

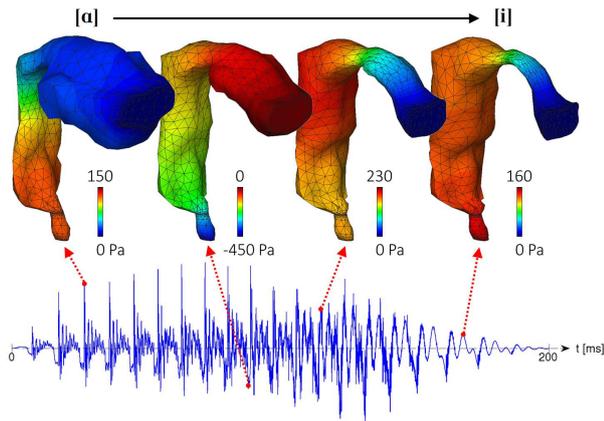


Figure 4: Snapshots from the FEM simulation of the vowel-vowel sequence [ai] showing the acoustic pressure values on the walls and the finite element mesh. Values were taken at different time instants, capturing the articulation of vowel [a] ($t=27$ ms) and vowel [i] ($t=168$ ms), and the transition between them ($t=88, 115$ ms). The acoustic pressure evolution tracked close to the mouth (generated sound) is also represented in the bottom of the figure with red dots indicating these time instants.

results. Very similar transfer functions can be observed for both approaches at frequencies below 5 kHz. Only some small discrepancies have been found for vowel [u] (see fourth formant), which might be reduced increasing the number of cross-sections. As far as the high frequency content is concerned, larger differences are observed for all the vowels, as this frequency range is more sensitive to small variations in the vocal tract shape (see e.g., [19, 28]). Again, these may be reduced by increasing the number of cross-sections. On the other hand, note that none of the vocal tract transfer functions decay in frequency as usually does the human voice, since the definition of $H(f)$ compensates somehow for this effect. The typical spectral tilt would appear for a proper train of glottal pulses introduced at the vocal tract entrance.

Despite the observed differences, these results show that the SPG approach can be used with confidence for 3D vowel simulations, recovering to a large extent most of the formants that the AG method produces.

3.2. Vowel-vowel sequence [ai]

The vowel-vowel sequence [ai] was simulated as an example of application of interpolation using the SPG. As explained in Section 2.2.3, we started from an initial finite element mesh corresponding to an intermediate vocal tract geometry between the articulation of [a] and [i]. This initial finite element mesh was first moved to vowel [a]. Then, acoustic wave propagation through the dynamic vocal tract was simulated, capturing the acoustic pressure at a node close to the mouth exit so as to listen to the generated vowel-vowel sequence [ai]. The articulation of vowel [a] and [i] were respectively sustained for 15 ms and 35 ms, while the transition time between both vowels was set to 150 ms. Linear interpolation was used between the two geometries. This gives a total time of 200 ms for the production of [ai].

Figure 4 shows four snapshots corresponding to different time instants of the FEM simulation of the vowel-vowel sequence [ai]. The acoustic pressure distribution on the vocal tract

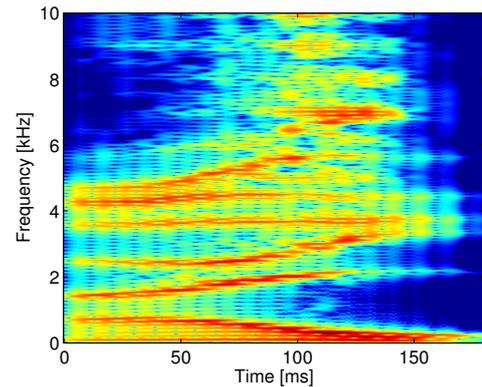


Figure 5: Spectrogram of the simulated vowel-vowel sequence [ai]. A pre-emphasis filter was used to enhance the visualization of higher frequencies.

walls can be observed together with the finite element mesh. The color scale was adapted for each time instant to enhance the visualization of the acoustic waves. The first snapshot starting from the left corresponds to the articulation of vowel [a], the latter to vowel [i], while intermediate ones correspond to the transition between the two vowels. These four snapshots were respectively captured at time instants $t=(27, 88, 115, 168)$ ms. These time instants are also represented as red dots in the bottom of Figure 4, which shows the acoustic pressure evolution of the generated vowel-vowel sequence. Figure 5 shows the corresponding spectrogram. As it can be observed, the formants smoothly transition from those of vowel [a] to those of vowel [i].

4. Conclusions

The semi-polar grid (SPG) was shown to be a valid tool for the 3D simulation of vowel-vowel sequences, as it simplifies a 3D interpolation problem to a 2D problem where only cross-sections have to be interpolated. Moreover, the SPG helps preventing the intersection of cross-section as their orientations are fixed over time. For vowels [a], [i] and [u] the SPG provided very similar results to those of the adaptive grid (AG) method, which extracts the cross-sections perpendicular to the vocal tract midline. As an example of application, the vowel-vowel sequence [ai] was generated using the SPG, resulting in a smooth transition of the formants. Future work will involve the comparison of 3D results against 1D, using both the SPG and AG approaches. More vowel-vowel utterances will also be generated.

5. Acknowledgements

This research has partially been supported by EU-FET grant EUNISON 308874. The first and third authors also acknowledge the Agencia Estatal de Investigación (AEI) and FEDER, EU, through project GENIOVOX TEC2016-81107-P, the grant 2014-SGR-0590 from the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (Generalitat de Catalunya), and the support of grants 2016-URL-IR-010 and 2016-URL-IR-013 from the Generalitat de Catalunya and the Universitat Ramon Llull.

6. References

- [1] P. Švancara and J. Horáček, “Numerical modelling of effect of tonsillectomy on production of Czech vowels,” *Acta Acustica united with Acustica*, vol. 92, no. 5, pp. 681–688, 2006.
- [2] T. Vampola, J. Horáček, and J. G. Švec, “FE modeling of human vocal tract acoustics. Part I: Production of Czech vowels,” *Acta Acustica united with Acustica*, vol. 94, no. 5, pp. 433–447, 2008.
- [3] M. Arnela and O. Guasch, “Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method,” *Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 4197–4209, 2013.
- [4] M. Arnela, O. Guasch, and F. Aliás, “Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations,” *Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 2946–2954, 2013.
- [5] H. Matsuzaki, N. Miki, and Y. Ogawa, “3D finite element analysis of japanese vowels in elliptical sound tube model,” *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 83, no. 4, pp. 43–51, 2000.
- [6] T. Kako and K. Touda, “Numerical method for voice generation problem based on finite element method,” *Journal of Computational Acoustics*, vol. 14, no. 1, pp. 45–56, 2006.
- [7] A. Hannukainen, T. Lukkari, J. Malinen, and P. Palo, “Vowel formants from the wave equation,” *Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. EL1–EL7, 2007.
- [8] H. Takemoto, P. Mokhtari, and T. Kitamura, “Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method,” *Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3724–3738, 2010.
- [9] H. Takemoto, S. Adachi, P. Mokhtari, and T. Kitamura, “Acoustic interaction between the right and left piriform fossae in generating spectral dips,” *Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 2955–2964, 2013.
- [10] H. Takemoto, P. Mokhtari, and T. Kitamura, “Comparison of vocal tract transfer functions calculated using one-dimensional and three-dimensional acoustic simulation methods,” in *INTERSPEECH 2014 – 15th Annual Conference of the International Speech Communication Association, September 14-18, Singapore, Proceedings*, 2014, pp. 408–412.
- [11] R. Blandin, M. Arnela, R. Laboissière, X. Pelorson, O. Guasch, A. Van Hirtum, and X. Labal, “Effects of higher order propagation modes in vocal tract like geometries,” *Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 832–843, 2015.
- [12] R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, “Influence of higher order acoustical propagation modes on variable section waveguide directivity: Application to vowel [a],” *Acta Acustica united with Acustica*, vol. 102, no. 5, pp. 918–929, 2016.
- [13] M. Speed, D. T. Murphy, and D. M. Howard, “Three-dimensional digital waveguide mesh simulation of cylindrical vocal tract analogs,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 449–455, 2013.
- [14] M. Speed, D. Murphy, and D. Howard, “Modeling the vocal tract transfer function using a 3d digital waveguide mesh,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 453–464, 2014.
- [15] D. Aalto, O. Aaltonen, R.-P. Happonen, P. Jääsaari, A. Kivelä, J. Kuortti, J.-M. Luukinen, J. Malinen, T. Murtola, R. Parkkola, J. Saunavaara, T. T. Soukka, and M. Vainio, “Large scale data acquisition of simultaneous MRI and speech,” *Applied Acoustics*, vol. 83, pp. 64–75, 2014.
- [16] S. Dabbaghchian, M. Arnela, O. Engwall, O. Guasch, I. Stavness, and P. Badin, “Using a biomechanical model and articulatory data for the numerical production of vowels,” in *INTERSPEECH 2016 – 17th Annual Conference of the International Speech Communication Association, September 8-12, San Francisco, USA, Proceedings*, 2016, pp. 3569–3573.
- [17] M. Arnela, S. Dabbaghchian, O. Guasch, and O. Engwall, “Finite element generation of vowel sounds using dynamic complex three-dimensional vocal tracts,” in *23th International Congress on Sound and Vibration (ICSV23), July 10-14, Athens, Greece, Proceedings*, 2016, pp. 1395–1402.
- [18] S. Dabbaghchian, M. Arnela, and O. Engwall, “Simplification of vocal tract shapes with different levels of detail,” in *18th International Congress of Phonetic Sciences (ICPhS), August 10-14, Glasgow, Scotland, UK, Proceedings*, 2015, pp. 1–5.
- [19] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, A. Van Hirtum, and X. Pelorson, “Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds,” *Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. 1707–1718, 2016.
- [20] B. H. Story, “Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002,” *Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 327–335, 2008.
- [21] —, “Phrase-level speech simulation with an airway modulation model of speech production,” *Computer Speech and Language*, vol. 27, no. 4, pp. 989–1010, 2013.
- [22] P. Birkholz, “Modeling consonant-vowel coarticulation for articulatory speech synthesis,” *PLoS ONE*, vol. 8, no. 4, p. e60603. doi:10.1371/journal.pone.0060603, 2013.
- [23] M. Arnela, O. Guasch, R. Codina, and H. Espinoza, “Finite element computation of diphthong sounds using tuned two-dimensional vocal tracts,” in *7th Forum Acousticum, September 7-12, Kraków, Poland, Proceedings*, 2014, pp. 1–6.
- [24] O. Guasch, M. Arnela, R. Codina, and H. Espinoza, “A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production,” *Acta Acustica united with Acustica*, vol. 102, no. 1, pp. 94–106, 2016.
- [25] M. Arnela, R. Blandin, S. Dabbaghchian, O. Guasch, F. Aliás, X. Pelorson, A. Van Hirtum, and O. Engwall, “Influence of lips on the production of vowels based on finite element simulations and experiments,” *Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2852–2859, 2016.
- [26] J. Cai, Y. Laprie, J. Busset, and F. Hirsch, “Articulatory modeling based on semi-polar coordinates and guided PCA technique,” in *INTERSPEECH 2009 – 10th Annual Conference of the International Speech Communication Association, September 6-10, Brighton, United Kingdom, Proceedings*, 2009, pp. 56–59.
- [27] A. E. Rosenberg, “Effect of glottal pulse shape on the quality of natural vowels,” *Journal of the Acoustical Society of America*, vol. 49, no. 2, pp. 583–590, 1971.
- [28] K. Motoki, “Three-dimensional acoustic field in vocal-tract,” *Acoustical Science and Technology*, vol. 23, no. 4, pp. 207–212, 2002.