



# Synthesis of VV Utterances from Muscle Activation to Sound with a 3D Model

Saeed Dabbaghchian<sup>1</sup>, Marc Arnela<sup>2</sup>, Olov Engwall<sup>1</sup>, Oriol Guasch<sup>2</sup>

<sup>1</sup>Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Sweden

<sup>2</sup>GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Spain

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

## Abstract

We propose a method to automatically generate deformable 3D vocal tract geometries from the surrounding structures in a biomechanical model. This allows us to couple 3D biomechanics and acoustics simulations. The basis of the simulations is muscle activation trajectories in the biomechanical model, which move the articulators to the desired articulatory positions. The muscle activation trajectories for a vowel-vowel utterance are here defined through interpolation between the determined activations of the start and end vowel. The resulting articulatory trajectories of flesh points on the tongue surface and jaw are similar to corresponding trajectories measured using Electromagnetic Articulography, hence corroborating the validity of interpolating muscle activation. At each time step in the articulatory transition, a 3D vocal tract tube is created through a cavity extraction method based on first slicing the geometry of the articulators with a semi-polar grid to extract the vocal tract contour in each plane and then reconstructing the vocal tract through a smoothed 3D mesh-generation using the extracted contours. A finite element method applied to these changing 3D geometries simulates the acoustic wave propagation. We present the resulting acoustic pressure changes on the vocal tract boundary and the formant transitions for the utterance [ai].

**Index Terms:** speech production, air-tight geometry, Finite Element Method, biomechanical model, acoustic model, deformable vocal tract, vowel-vowel sequences

## 1. Introduction

The human speech production apparatus is a complex, dynamic system in 3D, involving several different articulators and a large number of muscles controlling them. However, to reduce the complexity in simulations, these often make simplifications, such as synthesizing sound by interpolation of the area functions [1], two-dimensional vocal tracts [2] or 3D tubular vocal tracts [3, 4], of the constituent phonemes. When using area functions, one assumes plane wave propagation, which is only accurate for frequencies below 4-5 kHz [5]. For higher frequencies, more realistic 3D geometries of the vocal tract are required. Increase in computational power and the availability of numerical methods now make it possible to simulate the propagation of the acoustic waves in 3D.

However, although interpolation of 3D vocal tract tubes provide great insights into speech acoustics, they may not reflect the dynamic of the speech production properly. An alternative to these approaches are methods using articulatory models, which can be either geometrical [6, 7, 8] or biomechanical [9, 10, 11]. Geometrical models are developed to replicate the kinematic of the articulation rather than its dynamics, while biomechanical models are more suitable for reproducing the speech dynamics.

It seems that there are three reasons for using area functions

for synthesis. First, traditional 1D acoustic models are well developed for the area function. Second, area function is estimated from the vocal tract midsagittal contour of 2D midsagittal articulatory models [9]. Third, constructing the vocal tract 3D geometry is not as simple as calculating the area function in an articulatory model and especially not in a biomechanical model [12]. The first and second reasons are not valid anymore, since 3D articulatory biomechanical models [10, 11] and acoustic models [13, 14] are currently available. However, construction of a 3D vocal tract geometry, which is a requirement for the acoustic simulation, is still a challenge in particular when using articulatory models that include separate articulators. That is, such a geometry requires a labor work by an expert before it can be used for the acoustic simulation. This is not a large problem when dealing with one or two geometries as in the static case. However, as an example, we need 300 geometries for a diphone simulation with 300 ms duration and 1 kHz sampling rate. Obviously, it is not reasonable to generate 300 geometries manually.

In our previous work, we developed an automatic method to generate an air-tight vocal tract geometry from a 3D biomechanical model [12]. Then we applied this technique for sound generation of three cardinal vowels. This work is an extension of our previous work for the production of vowel-vowel sequences where the benefit of this automatic method is obvious. Furthermore, the acoustic model has been improved to deal with moving boundaries of the vocal tract, which appear in sequences of sounds. By coupling biomechanical and acoustic models, production of [ai] is simulated from contraction of muscles to sound.

## 2. Methods

Simulation of a diphone sound involves biomechanical model, vocal tract geometry construction, and acoustic model which are explained in this section.

### 2.1. Biomechanical model

We use ArtiSynth ([www.artisynth.org](http://www.artisynth.org)) for biomechanical modeling [15]. An improved version of a 3D biomechanical model, which was developed in our previous work [12], is used in this study. The lips, which were missing, were added to the model. The upper and lower lips were cropped from the face model and attached to the maxilla and mandible respectively. This means that the upper lip is fixed and the lower lip follows the jaw movement. Since the face muscles were excluded, there is no control for the lips aperture/protrusion. This should not be a problem for [ai] where both constituent vowels are unrounded (see [16] for the acoustic influence of the lips). Figure 1a depicts a snapshot of the biomechanical model. In this model the tongue, jaw, lower lip, and hyoid bone are dynamic. Other structures including the upper lip, maxilla, soft palate, pharynx wall, and larynx are static. A list of the jaw and tongue muscles are pre-

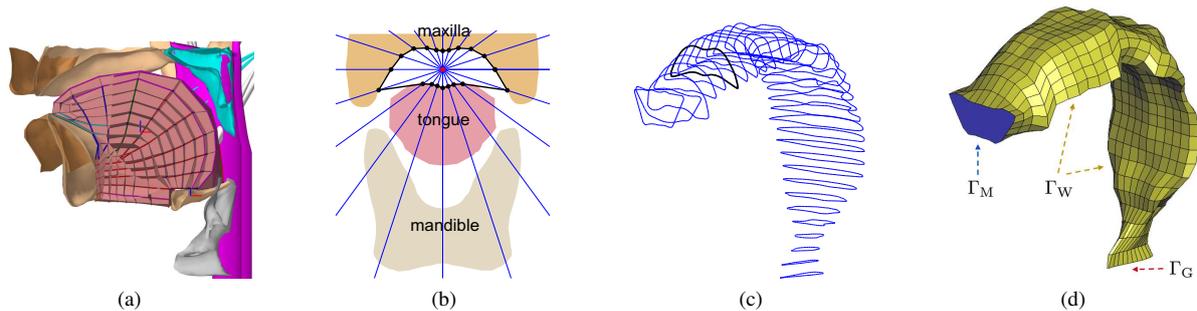


Figure 1: (a) A snapshot of 3D biomechanical model developed in ArtiSynth, (b) an example of coronal cross-section in the oral cavity and reconstruction of the vocal tract contour, (c) cross-sections of the reconstructed vocal tract, (d) initial geometry with boundaries  $\Gamma_G$ ,  $\Gamma_W$  and  $\Gamma_M$  respectively denoting the glottis, the vocal tract walls and the mouth aperture.

sented in Table 1. For more details about the model, we refer reader to the previous publications [11, 12, 17, 18].

## 2.2. Vocal tract geometry

The human vocal tract is a cavity rather than a physical object. When we have defined articulators such as tongue, palate, pharynx, etc., the vocal tract has already been defined implicitly as a cavity enclosed by all articulators. Because of that it is not reasonable to define an explicit geometry of the vocal tract in a biomechanical model. Based on this idea, we proposed an approach to extract an air-tight vocal tract 3D geometry in our previous work [12]. This approach has been improved regarding robustness, and computational efficiency in this work to be applicable for changing sound.

A set of gridplanes, arranged in a semi-polar grid are defined [19]. Then, 3D geometries (i.e. surface meshes) of the articulators are intersected with these gridplanes. This results in several planar cross-sections. On each cross-section, there are several contours, each corresponding to an articulator. Figure 1b exemplify a coronal cross-section in the oral cavity with three contours of the maxilla, mandible and tongue. These cross-sections are processed to determine the cavity or airway contour. In an ideal situation, we may expect to see a hole enclosed by several contours. What we see instead is several contours with gaps between them, and no such a hole in Figure 1b. To extract the hole, a center point is calculated as shown with the red dot in Figure 1b. Then a set of rays starting from this center and spreading the whole range of angles between 0 and  $2\pi$  is defined. The intersection of these rays and the contours of the articulators define the outline of the vocal tract cross-section. The process is repeated for all of the cross-sections. Finally, all hole contours are projected back into 3D space (see Figure 1c) and a surface is wrapped to construct the air-tight geometry as shown in Figure 1d. Using this method, the main branch of the cavity is reconstructed, but the side branches, such as sublingual cavity in our example in Figure 1b, are omitted. However, the method is capable to include side branches with some improvements. For each side branch, we need a center point to grow a region around that.

The original extracted geometries are not smooth enough and may cause problem for acoustic simulations (i.e. volume mesh generation). Smoothness in spatial domain is not enough for a diphone production and it is important to have smooth geometries over time (i.e. smooth velocity of the vertices). Thus, extracted geometries are smoothed in both spatial and time domains in a post processing step. We use a combination of

Laplacian smoothing and spline curves for spatial and temporal processing respectively. This post processing is extremely important to adjust the smoothing degree in order not to affect the constriction area significantly.

## 2.3. Acoustic model

The generation of a diphone involves acoustic waves propagating through a dynamic vocal tract. This physical phenomenon can be described by the mixed wave equation for the acoustic pressure  $p(\mathbf{x}, t)$  and acoustic particle velocity  $\mathbf{u}(\mathbf{x}, t)$ , expressed in an ALE (Arbitrary Lagrangian-Eulerian) frame of reference to account for the vocal tract movement [3]. The equation reads

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

$$\rho_0 \partial_t \mathbf{u} - \rho_0 \mathbf{u}_{\text{dom}} \cdot \nabla \mathbf{u} + \nabla p = 0, \quad (1b)$$

with  $c_0$  standing for the speed of sound,  $\rho_0$  for the air density,  $\mathbf{u}_{\text{dom}}$  for the velocity of the domain (i.e. the vocal tract), and  $\partial_t$  for the first partial time derivative. The ALE mixed wave equation (1) is supplemented with the following boundary and initial conditions,

$$\mathbf{u} \cdot \mathbf{n} = g(t) \quad \text{on } \Gamma_G, t > 0, \quad (2a)$$

$$\mathbf{u} \cdot \mathbf{n} = p/Z_w \quad \text{on } \Gamma_W, t > 0, \quad (2b)$$

$$p = 0 \quad \text{on } \Gamma_M, t > 0, \quad (2c)$$

$$p = 0, \mathbf{u} = 0 \quad \text{in } \Omega, t = 0, \quad (2d)$$

with  $\Omega$  standing for the computational domain with boundaries  $\Gamma_G$ ,  $\Gamma_W$  and  $\Gamma_M$  (see Figure 1d). In Eq. (2a) a particle velocity  $g(t)$  is introduced at the the fictitious boundary  $\Gamma_G$  where the glottis is located. In Eq. (2b) the impedance  $Z_w$  is used to account for losses on the vocal tract walls  $\Gamma_W$ , and in Eq. (2c) an open-end boundary condition is prescribed on the mouth aperture  $\Gamma_M$ . The latter boundary condition does not consider radiation losses. These could be implemented by extending the computational domain outside the vocal tract [20, 21], but at the prize of increasing the computational cost.

The sequence [ai] was generated by numerically solving the ALE mixed wave equation (1) with boundary and initial conditions (2) using the Finite Element Method (FEM). In particular, a subgrid scale strategy was followed to avoid numerical instabilities when using the same interpolation for the acoustic pressure and the acoustic particle velocity. Details of the numerical implementation can be found in [3].

Table 1: Percentage of estimated activation of tongue and jaw muscles for two vowels. Zero and hundred represent no active and maximum active force generated by the muscle respectively.

	GGP	GGM	GGA	SG	HG	MH	GH	V	T	IL	SL	JO	JC
a	0	5	20	0	15	0	22	7	0	0	0	0.02	0
i	100	0	0	0	3	15	0	0	58	63	19	0	1.05

Tongue muscles, *GGP*: genioglossus posterior, *GGM*: genioglossus middle, *GGA*: genioglossus anterior, *SG*: styloglossus, *HG*: hyoglossus, *GH*: geniohyoid, *MH*: mylohyoid, *V*: verticalis, *T*: transversus, *IL*: inferior longitudinal, *SL*: superior longitudinal. Jaw muscles, *JO*: bilateral opener, *JC*: bilateral closer.

In the simulations the speed of sound was set to  $c_0 = 350$  m/s and the wall impedance to  $Z_w = 83666$  kg/m<sup>2</sup>s [22]. A train of glottal pulses was generated using a Rosenberg model [23] to obtain the signal  $g(t)$ . These glottal pulses were enhanced by considering a pitch curve, a fade in/out to emulate the onset/offset of the vocal folds, and some shimmer and jitter. The sampling frequency used to discretize the computation time interval  $[0, T]$  in constant time steps  $\Delta t$  was set to  $f_s = 1/\Delta t = 200$  kHz.

A volumetric finite element mesh moving with the vocal tract geometry was also needed for the simulation. We started from an initial finite element mesh with tetrahedral elements of size 0.004 m, which was generated from a vocal tract geometry corresponding to an articulation between vowels [a] and [i] (see Figure 1d). This mesh was deformed to reach the articulation of vowel [a], and then to vowel [i] so as to generate the sequence [ai]. This procedure allows one to minimize element distortion and to avoid remeshing strategies which are very time consuming.

The coordinates of the boundary nodes located on the vocal tract walls  $\mathbf{x}_{\text{walls}}(t)$  were prescribed according to the reconstructed geometries from the biomechanical model. However, one also has to find the inner node positions of the volumetric mesh at each time step  $t^n$ . To do so, an additional problem was solved using FEM. This consisted in the resolution of the Laplacian equation for the node displacements  $\mathbf{w}$ ,

$$\nabla^2 \mathbf{w}^{n+1} = 0 \quad \text{in } \Omega, t = t^{n+1}, \quad (3a)$$

with boundary conditions

$$\mathbf{w}^{n+1} = \mathbf{x}_{\text{walls}}^{n+1} - \mathbf{x}_{\text{walls}}^n \quad \text{on } \Gamma_W, t = t^{n+1}, \quad (3b)$$

$$\mathbf{w}^{n+1} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_G, t = t^{n+1}, \quad (3c)$$

$$\mathbf{w}^{n+1} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_M, t = t^{n+1}. \quad (3d)$$

This allows one to smoothly translate the movement of the boundary nodes to the inner nodes. Note that in Eq. (3b) the wall movement is prescribed, while in Eq. (3c) and Eq. (3d) the displacement is set to zero in the normal direction  $\mathbf{n}$  (pointing outwards) to avoid an artificial lengthening of the vocal tract.

### 3. Results and Discussion

The articulatory transition was generated through activation of muscles in the biomechanical model. We here report on the articulatory and acoustic outcome of this activation.

#### 3.1. Activation of muscles

Activity of a muscle is proportional to its active force generation and is represented with a number in the range  $[0, 1]$ , where zero

and one indicate no and maximum active force generation respectively. Using inverse modeling and EMA (Electromagnetic Articulography) data, activation of jaw and tongue muscles was estimated for vowel [a] and [i] as proposed in [12]. However, to get more reasonable estimation, some of the muscles, depending on the vowel, were excluded to be used by the inverse simulation [24]. The results are presented in Table 1. Based on these activations, a temporal trajectory was generated by using linear interpolation between these targets as shown in Figure 2. In this figure, the values of the y-axis are scaled non-linearly to show wide range of the muscles activation. Reaching a target value, activations remain constant for several milliseconds. This time is shorter for the first target [a] than the second target [i]. The reason is that muscle activations for [a] are smaller than [i] and so it needs shorter duration before reaching equilibrium.

By linear interpolation between target activations, we have implicitly assumed that the speech production is planned in muscle activation space. However, regardless of if this assumption is correct or not, it does not affect the core of this work.

#### 3.2. Articulation

To investigate the articulation of the tongue and jaw, one point on the jaw, and three points on the tongue surface (tip, middle, and back) were marked and their spatial trajectory was captured. Figure 3a shows the articulators contour at  $t = 0$  and the trajectories of these four marked points in the midsagittal plane. These trajectories are very similar to trajectories captured by Electromagnetic Articulography [25]. This is interesting because we observe that only a simple linear interpolation between two targets of muscles activation can generate such non-linear trajectories. A similar results has been reported in [26] using 2D biomechanical model where the authors concluded that this kind of trajectory is a result of biomechanical constraints rather than control mechanism.

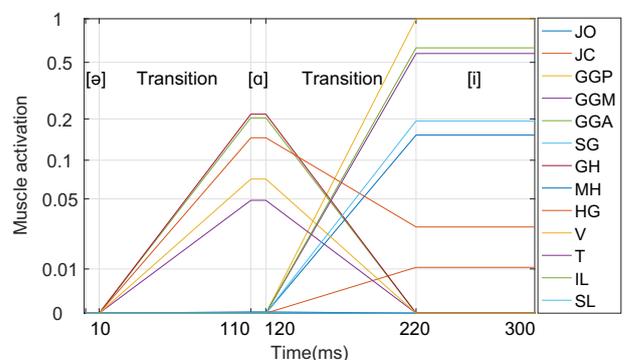


Figure 2: Trajectory of muscles activation

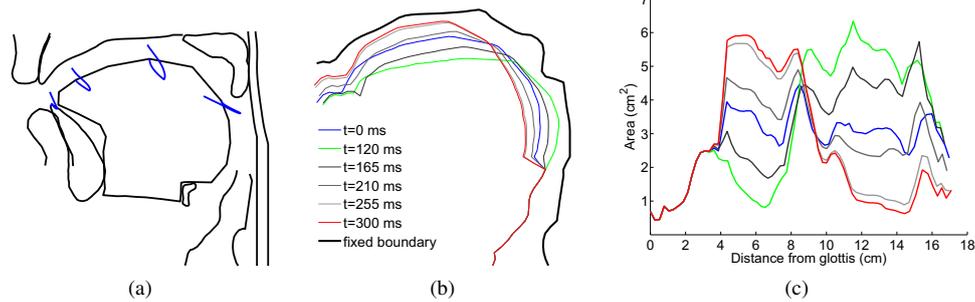


Figure 3: (a) Trajectories of jaw and tongue markers, (b) deformation of the vocal tract geometry in the midsagittal plane, and (c) corresponding area functions.

Figure 3b illustrates the temporal deformation of the vocal tract in the midsagittal plane at different times during the movement from neutral ( $t = 0$ ) to vowel [a] ( $t = 110$ ) and vowel [i] ( $t = 220$ ). The corresponding area functions are also depicted in Figure 3c to illustrate the fact that the change in the cross-sectional areas is not linear over time. These area functions were acquired from the 3D vocal tract geometry employing the approach proposed in [14, 19]. In addition, Figure 4 visualizes the deformation of the vocal tract 3D geometry during the production of [ai].

### 3.3. Acoustics

Figure 4 presents four snapshots acquired during the FEM simulation of [ai] at time instants 130, 190, 230 and 270 ms. The figure shows the acoustic pressure distribution on the vocal tract walls together with the finite element mesh on the boundaries. The color scale was adapted to each frame to enhance the visualization of acoustic waves.

Besides, the acoustic pressure evolution was also tracked on a node close to the mouth exit. Once converted to an audio file one can listen to the generated sound for [ai] (see supplementary audio file).

The spectrogram of [ai] is shown in Figure 5. As usually done in speech analysis, a pre-emphasis filter was used to enhance the visualization of the high frequency range. It can be observed how the typical formants of vowel [a] smoothly transit to those of vowel [i] during the production of the sequence.

## 4. Conclusions

An automatic method was proposed in this paper to generate deformable air-tight geometries of the vocal tract using the geometries of the surrounding articulators. Furthermore, an acoustic model was developed to solve the wave equation in time domain which deals with moving boundaries of the vocal tract. Simulation of a sequence [ai] from contraction of muscles to sound shows promising results. The proposed modeling approach in this work is important for several reasons. First, it creates the possibility to study speech production from different perspectives including physiology, articulation and acoustics. Second, 3D models are more realistic than using area functions or 2D models. For example, transversus muscle of the tongue, with lateral fiber directions, can not be modeled in 2D, and 3D acoustic simulation can model the high frequencies as well as low frequencies. Third, it sheds light on the dynamic aspects of the speech production whereas studying this in humans is still a challenge. Technical progress in real time MRI may help to cap-

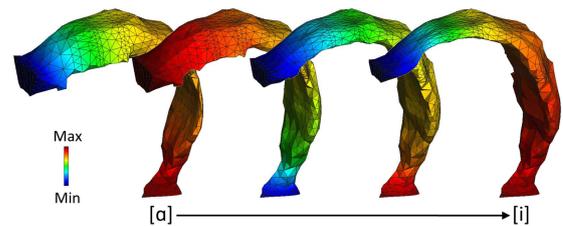


Figure 4: Snapshots of the FEM mesh and acoustic pressure distribution on the vocal tract wall for the sequence [ai].

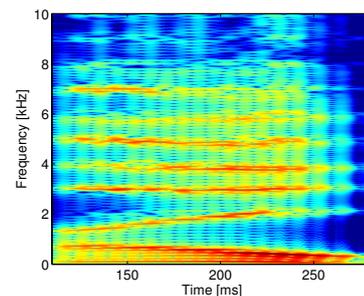


Figure 5: Spectrogram of the simulated sound [ai].

ture the movement of the vocal tract with high spatial and temporal resolution which leads to more knowledge about speech dynamics. Such crucial knowledge can be used to refine the model.

As future work, we would like to simulate more sequences and compare the dynamics of speech production between the geometries generated using the biomechanical model and the ones which are generated by interpolation between two static geometries of the vocal tract.

## 5. Acknowledgements

This research has partially been supported by EU-FET grant EUNISON 308874. The second and fourth 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] B. H. Story, "A parametric model of the vocal tract area function for vowel and consonant simulation," *The Journal of the Acoustical Society of America*, vol. 117, no. 5, p. 3231, 2005.
- [2] M. Arnela and O. Guasch, "Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 369–379, 2014.
- [3] 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.
- [4] M. Arnela, S. Dabbaghchian, O. Guasch, and O. Engwall, "Finite element generation of vowel sounds using dynamic complex three-dimensional vocal tracts," in *ICSV23 – 23th International Congress on Sound and Vibration*, Athens, Greece, 2016, pp. 1–7.
- [5] R. Blandin, M. Arnela, R. Laboissière, X. Pelorson, O. Guasch, A. Van Hirtum, and X. Laval, "Effects of higher order propagation modes in vocal tract like geometries," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 832–843, 2015.
- [6] O. Engwall, "Combining MRI, EMA and EPG measurements in a three-dimensional tongue model," *Speech Communication*, vol. 41, no. 2, pp. 303–329, 2003.
- [7] P. Birkholz, D. Jackel, and B. Kroger, "Construction And Control Of A Three-Dimensional Vocal Tract Model," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 1, 2006, pp. 1873–1876.
- [8] B. H. Story, "TubeTalker: An airway modulation model of human sound production," in *First International Workshop on Performative Speech and Singing Synthesis*, Vancouver, Canada, 2011.
- [9] Y. Payan and P. Perrier, "Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis," *Speech Communication*, vol. 22, pp. 185–205, 1997.
- [10] S. Buchaillard, P. Perrier, and Y. Payan, "A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning," *The Journal of the Acoustical Society of America*, vol. 126, no. 4, pp. 2033–2051, 2009.
- [11] P. Anderson, N. M. Harandi, S. Moisik, I. Stavness, and S. Fels, "A Comprehensive 3D Biomechanically-Driven Vocal Tract Model Including Inverse Dynamics for Speech Research," in *Interspeech 2015 – 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, 2015, pp. 2395–2399.
- [12] 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*, San Francisco, USA, 2016, pp. 3569–3573.
- [13] M. Arnela, O. Guasch, and F. Alías, "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations," *The Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 2946–2954, 2013.
- [14] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, A. V. Hirtum, and X. Pelorson, "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," *The Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. 1707–1718, 2016.
- [15] J. E. Lloyd, I. Stavness, and S. Fels, "ArtiSynth: a fast interactive biomechanical modeling toolkit combining multibody and finite element simulation," in *Soft tissue biomechanical modeling for computer assisted surgery*. Springer Berlin Heidelberg, 2012, pp. 355–394.
- [16] M. Arnela, R. Blandin, S. Dabbaghchian, O. Guasch, F. Alías, X. Pelorson, A. Van Hirtum, and O. Engwall, "Influence of lips on the production of vowels based on finite element simulations and experiments," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2852–2859, 2016.
- [17] I. K. Stavness, "Byte your tongue: A computational model of human mandibular-lingual biomechanics for biomedical applications," Ph.D. Thesis, University of British Columbia, 2010.
- [18] I. Stavness, J. E. Lloyd, Y. Payan, and S. Fels, "Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 27, no. 3, pp. 367–390, 2011.
- [19] S. Dabbaghchian, M. Arnela, and O. Engwall, "Simplification of vocal tract shapes with different levels of detail," in *ICPhS2015 – 18th International Congress of Phonetic Science*, Glasgow, Scotland, UK, 2015, pp. 1–5.
- [20] H. Takemoto, P. Mokhtari, and T. Kitamura, "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," *The Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3724–38, 2010.
- [21] M. Arnela and O. Guasch, "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 4197–4209, 2013.
- [22] 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.
- [23] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *The Journal of the Acoustical Society of America*, vol. 49, no. 2, pp. 583–590, 1971.
- [24] I. Stavness, J. E. Lloyd, and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model," *Journal of biomechanics*, vol. 45, no. 16, pp. 2841–8, 2012.
- [25] A. B. Youssef, "Control of talking heads by acoustic-to-articulatory inversion for language learning and rehabilitation," Ph.D. dissertation, Université de Grenoble, 2011.
- [26] P. Perrier, Y. Payan, M. Zandipour, and J. Perkell, "Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study," *The Journal of the Acoustical Society of America*, vol. 114, no. 3, pp. 1582–1599, 2003.