



HAL
open science

Using a biomechanical model and articulatory data for the numerical production of vowels

Saeed Dabbaghchian, Marc Arnela, Olov Engwall, Oriol Guasch, Ian Stavness, Pierre Badin

► **To cite this version:**

Saeed Dabbaghchian, Marc Arnela, Olov Engwall, Oriol Guasch, Ian Stavness, et al.. Using a biomechanical model and articulatory data for the numerical production of vowels. Interspeech 2016 - 17th Annual Conference of the International Speech Communication Association, Sep 2016, San Francisco, United States. pp.3569-3573, 10.21437/Interspeech.2016-1500 . hal-01368260

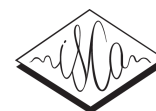
HAL Id: hal-01368260

<https://hal.science/hal-01368260>

Submitted on 19 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Using a Biomechanical Model and Articulatory Data for the Numerical Production of Vowels

Saeed Dabbaghchian¹, Marc Arnela², Olov Engwall¹, Oriol Guasch², Ian Stavness³, Pierre Badin⁴

¹Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Sweden

²GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Spain

³Department of Computer Science, University of Saskatchewan, Canada

⁴GIPSA-Lab, CNRS – Université Grenoble Alpes, France

saeedd@kth.se, marnela@salleurl.edu, engwall@kth.se, oguasch@salleurl.edu,
ian.stavness@usask.ca, Pierre.Badin@gipsa-lab.grenoble-inp.fr

Abstract

We introduce a framework to study speech production using a biomechanical model of the human vocal tract, ArtiSynth. Electromagnetic articulography data was used as input to an inverse tracking simulation that estimates muscle activations to generate 3D jaw and tongue postures corresponding to the target articulator positions. For acoustic simulations, the vocal tract geometry is needed, but since the vocal tract is a cavity rather than a physical object, its geometry does not explicitly exist in a biomechanical model. A fully-automatic method to extract the 3D geometry (surface mesh) of the vocal tract by blending geometries of the relevant articulators has therefore been developed. This automatic extraction procedure is essential, since a method with manual intervention is not feasible for large numbers of simulations or for generation of dynamic sounds, such as diphthongs. We then simulated the vocal tract acoustics by using the Finite Element Method (FEM). This requires a high quality vocal tract mesh without irregular geometry or self-intersections. We demonstrate that the framework is applicable to acoustic FEM simulations of a wide range of vocal tract deformations. In particular we present results for cardinal vowel production, with muscle activations, vocal tract geometry, and acoustic simulations.

Index Terms: speech production, biomechanical articulatory model, vocal tract geometry, vocal tract acoustics, Finite Element Method

1. Introduction

Human vocal tract models may be categorized into geometrical and biomechanical models. In a geometrical model, the vocal tract is represented by its initial geometry and a set of parameters directly deforms this geometry. Maeda [1] created a 2D midsagittal vocal tract model with seven control parameters, whereas Badin [2], Engwall [3] and Birkholz [4] proposed 3D vocal tract models with nine, six, and twenty-three control parameters, respectively. Story [5] designed a model of vocal tract area function controlled by two parameters. Geometrical models are computationally efficient but their application is limited to the study of speech acoustics and articulation. To study neuromuscular or motor control of speech production, a biomechanical model of speech production is needed. Different biomechanical models have been introduced and improved during the years. Payan [6] presented a 2D biomechanical tongue model to synthesize vowel-vowel sequences; Gérard [7] used a 3D biomechanical

tongue model to study speech motor control and Buchaillard [8] used a 3D tongue biomechanical model for cardinal vowel production. Other 3D biomechanical models were developed by Wu [9], Fang [10] and Stavness [11] to study muscle activation of tongue and jaw. Most recently Anderson [12] introduced a comprehensive biomechanical model of oropharyngeal structures.

Once a biomechanical model exists, two questions may arise: how to control the model and how to use it for acoustic simulations. The first question is the same for geometrical models, but the solution is usually more complex for biomechanical models, since they usually have more parameters than geometrical ones and those parameters of physically-based. Toutios [13] proposed a method to control an articulatory model by using Electromagnetic Articulography (EMA) data. Wu [9] presented an approach to control physiological model by using MRI.

The second question, which is the focus of this work, has not been fully addressed in past literature. Acoustic simulations require vocal tract geometry, either as a 1D area function [14] or as a 3D tissue-air interface [15]. In a biomechanical model there is no explicit representation of the vocal tract, because the vocal tract is a cavity rather than a physical object. However, changes in muscle activity move the articulators and hence the vocal tract geometry is indirectly deformed.

Using a 2D biomechanical model with a 1D area function as the vocal tract representation is the most common approach used in previous studies [16]. In this approach the vocal tract area function is estimated with $\alpha\beta$ model by using the midsagittal vocal tract outline [17], [18]. Although the use of 3D biomechanical models is growing [8]–[10], [12], [19], they still use the 1D representation of the vocal tract (area function) for 1D acoustic simulations [8], [12]. However, this assumes plane wave propagation, which holds for frequencies below 4 kHz, but is not valid for high frequencies [20], [21]. Therefore, 3D acoustic simulations should be performed to overcome this limitation [21], [22]. In order to couple a 3D tissue-air interface to a 3D biomechanical model, Stavness [23] proposed a “skinning” technique whereby a virtual vocal tract tube is aligned to and deforms with the surrounding biomechanical structures. This technique requires an initial effort to register skin vertices with other structures. Furthermore, the skinning approach with large deformations can lead to severe stretching and sharp corners in the vocal tract mesh making it unsuitable for volume mesh generation.

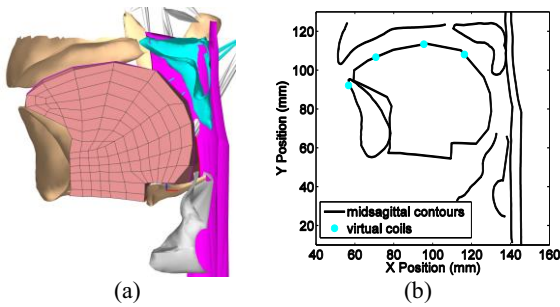


Figure 1: mid-sagittal cut of the biomechanical model (a), and position of virtual coils (b)

Another limitation of the skinning technique is that side branches of the vocal tract, such as the vallecula and sublingual cavity, are difficult to handle because they appear and disappear during speech production. In this paper, we propose an alternative, fully-automatic method to extract the vocal tract geometry and area function directly from the surrounding structures. Extracted geometries satisfy the mesh quality requirements for 3D FEM acoustic simulations and were used to generate speech sound by solving the wave equation in the time domain.

The possibility to automatically extract a vocal tract representation from a 3D biomechanical model, either as an area function or as a 3D geometry, makes it feasible to run large numbers of simulations of static sounds or dynamic sounds such as diphthongs. Our approach couples a 3D biomechanical model to 3D acoustic simulations, which is the main contribution of this work.

2. Methods

2.1. Biomechanical model

We have adapted a recently developed biomechanical model of the speech organs [12], [24] in ArtiSynth [25] for this study. The reason for the adaptation was that using the pharynx, soft palate and larynx from the original model [12], it was difficult to reach the right formant positions, especially for vowel [a]. There are several possible reasons for this difficulty. First, in EMA data used to estimate the tongue position, there is no coil on the pharyngeal part of the tongue (posterior one-third) where the constriction occurs for vowel [a]. Since the tongue model has many (~2500) degrees of freedom, prescribing positions or trajectories of three points on the tongue surface does not provide enough information to specify the whole tongue shape and position. Second, the inclination of the pharyngeal wall in the original model creates a large area in the oropharynx, even when the tongue is pulled back. We therefore adapted the pharynx wall to be more vertical than in the original model. Third, the larynx's relative position to tongue and the lack of an epiglottis in the model also contributes to a large area in the laryngopharynx. To alleviate this problem, the larynx was moved 7 mm towards the tongue. We also made the uvula a bit smaller to avoid its collision with the tongue. Figure 1a shows a mid-sagittal view of the model.

Estimating model parameters by considering the dynamics of all the articulators is very complex, and needs articulatory data, e.g. EMA data. In this study, the jaw and tongue are considered as dynamic articulators. The hard palate, soft palate, pharynx, and larynx are static and represented as surface meshes. The nasal cavity is closed by the soft palate.

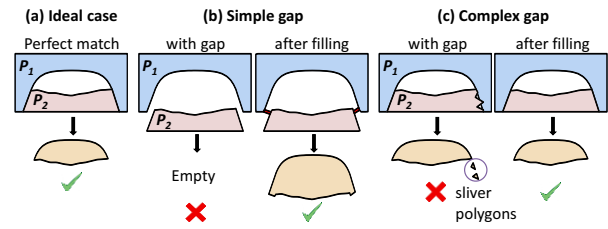


Figure 2: Illustration of the basic idea to extract an enclosed area between two polygons.

2.2. EMA data and inverse tracking

We use EMA data [26] for the same subject used for the biomechanical model of tongue, jaw, and hard palate [12]. The EMA material consisted of French vowels and three coils on the tongue, one coil on the jaw, and two coils on the upper and lower lips.

For each coil of the EMA data, a corresponding virtual coil is positioned as a marker in the model (Figure 1b) and a linear trajectory from rest to target coil position was used as input to the inverse tracking controller in ArtiSynth [11]. The controller solves for muscle activations at each simulation time step in order to drive the biomechanical model to achieve the target coil positions while also minimizing the sum of activations squared. The details of the tracking controller formulation have been previously reported [11], [25].

2.3. Vocal tract geometry

To form the boundary of the vocal tract, it is considered as a cavity surrounded by some other structures rather than an individual physical object. Figure 2 illustrates the general idea with two polygons in 2D space. When the adjacent borders of the two polygons match perfectly, applying Union operation to these polygons generates another polygon with a hole. This hole is equivalent to the enclosed area between the polygons (Figure 2a). This idea can be extended when there are more polygons. In reality, adjacent borders do not match perfectly and usually there is a gap or overlap between them. In the context of the biomechanical model, this mismatch is introduced in the model development and has several possible reasons, including segmentation errors of structures, registration of structures, inconsistent discretization of borders, and missing objects. Although this mismatch may be reduced by improving the techniques for model development, it is almost impossible to avoid them completely. Even if the boundaries of the model match perfectly at rest, moving the objects will introduce mismatches. Two possible mismatches occur, namely simple gaps and complex gaps. A simple gap is when two polygons are far apart and there is no overlap at all. In this case, there is no enclosed area between the two polygons. One or more filling polygons are placed to enclose the area (see Figure 2b, for an example). A complex gap occurs when two polygons intersect each other several times resulting in several enclosed areas, of which only the main one is of interest. In such cases, the adjacent borders with several intersections are detected and snapped together (Figure 2c). The whole idea is applicable to 3D space in exchange for more complexity and computational cost. In the context of our work, we convert the 3D problem to a set of 2D problems by sampling the 3D space with some planes; we have used a semi-polar grid [27], with 20 horizontal, 30 oblique, and 20 vertical gridplanes in different sections of the grid, for this

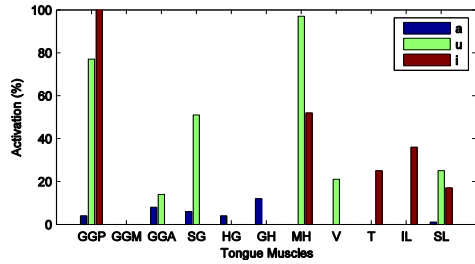


Figure 3: Estimated percentage of activation of the tongue extrinsic and intrinsic muscles

conversion. Since the semi-polar grid is nearly perpendicular to the vocal tract, it makes subsequent processing easier.

Since the lips are missing from the model, in order to get the right formants, especially for [u], a cylinder was attached to the mouth opening. The cylinder parameters (I_a, I_p) , where I_a corresponds to lips aperture in cm^2 and I_p corresponds to lips protrusion in cm) was chosen as $(4.5, 0.8)$ for [a], $(0.3, 1.5)$ for [u] and $(3, 0.5)$ for [i] following [8].

2.4. Acoustic simulations

Three dimensional sound wave propagation through the vocal tract was simulated using FEM by solving the time-domain wave equation for the acoustic pressure $p(x, t)$ [28],

$$\partial_{tt}^2 p - c_0^2 \nabla^2 p = 0. \quad (1)$$

In Eq. (1) c_0 stands for the speed of sound and ∂_{tt}^2 designates the second order derivative with respect to time. In what concerns boundary conditions, a wall impedance of $Z_w=83666 \text{ kg/m}^2\text{s}$ [29] was imposed at the vocal tract walls to account for boundary losses, a volume inflow $Q_g(t)$ was introduced at the glottal cross-section where the vocal cords are located, and a zero pressure release condition ($p = 0$) was imposed at the mouth sectional area to consider an open-end condition. Although radiation losses were not considered in this study, they could be introduced by allowing sound waves to propagate out from the vocal tract exit towards infinity (see e.g., [22]).

Each of the generated vowel vocal tract geometries were 3D meshed using tetrahedral elements of size $h \approx 0.003 \text{ m}$. Acoustic simulations were then conducted using a speed of sound of $c_0 = 350 \text{ m/s}$ and a sampling frequency of 250 kHz . Two kinds of simulations were performed. In the first one we analyzed the vocal tract acoustic response by computing the so-called vocal tract transfer function,

$$H(f) = P_o(f)/Q_g(f), \quad (2)$$

with $P_o(f)$ and $Q_g(f)$ respectively standing for the Fourier Transform of the acoustic pressure at the mouth and the volume velocity introduced at vocal tract entrance (glottal cross-section). In this case we used for $Q_g(t)$ a Gaussian pulse and we simulated a 50 ms event. In the second kind of

simulations we generated a vowel sound by introducing a train of glottal pulses of the Rosenberg type [30] at the glottal cross-section and collecting the evolution of the acoustic pressure at the vocal tract exit, 0.003 m from the mouth center.

3. Results

The proposed framework was used for the numerical production of the cardinal vowels [a], [i], and [u]. For each vowel, muscles activation, articulatory posture, vocal tract geometry, and acoustic simulations are presented.

3.1. Muscle activation

Figure 3 shows the estimated percentage of activation for the tongue muscles including genioglossus posterior (GGP), genioglossus middle (GGM), genioglossus anterior (GGA), styloglossus (SG), hyoglossus (HG), geniohyoid (GH), mylohyoid (MH), verticalis (V), transversus (T), inferior longitudinal (IL), and superior longitudinal (SL). Validation of the model estimation with experiment is still an open question, especially for intrinsic tongue muscles and there are very few studies where muscle activation was measured during speech production. Electromyography (EMG) technique is very invasive and not practical for measurements of speech production. Baer [31] reported the measurement of extrinsic tongue muscle during the production of vowels v in $[\text{a}pvv]$ context. In his work, GG is divided into two parts namely GGA (this corresponds to GGM and GGA in our model) and GGP. Wu [9] normalized the EMG measurements [31] for comparison. Figure 4 shows estimated muscle activation from the model and EMG measurements [9], [31] for the three cardinal vowels.

[a]: Estimated values are small compared to EMG data. Articulatory compensation [1], [32] could be a possible reason. It might be that to make a constriction in oropharynx, the study's speaker employs a different strategy compared to the speaker in [31]. With relatively large mouth opening by the study's speaker, there is no need for a large force to push the tongue back since it is already moved back by lowering the jaw.

[u]: both model and EMG data shows high level of activation for GGP and SG. According to EMG data, HG is active while the model estimated no activity. Since the GGA and GGM are close to each other, the model and the EMG data are nevertheless rather similar to each other.

[i]: both model and EMG shows saturation force of GGP. The model prediction for GGM and GGA is zero while almost 25% activation was reported in the measurement.

In another study by Honda [33], it has been reported that intrinsic muscles are not important for vowels and they are mainly responsible for tongue blade deformation in consonants. According to this study, HG and SG are the two

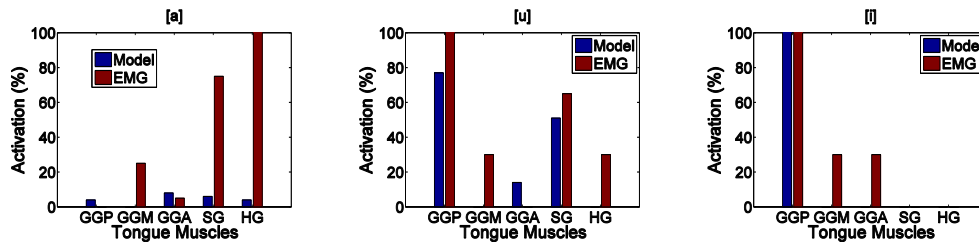


Figure 4: Estimated extrinsic tongue muscles and EMG measurements [9], [31]

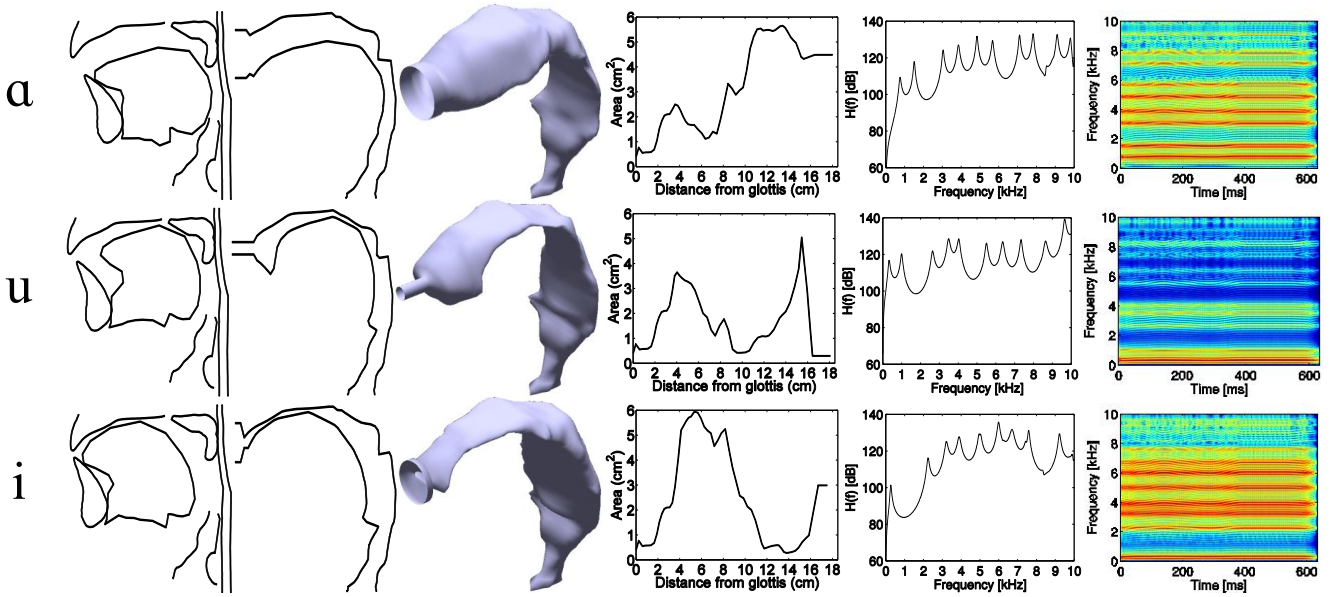


Figure 5: Articulators postures, midsagittal contour, 3D geometry, area function, transfer function of the vocal tract, and spectrogram of the generated sounds

main active muscles for vowel [a]; GGP and SG for [u]; GGP and GGA for [i]. Comparing this with our findings for extrinsic muscles, the overall results seem to be reasonable. One inconsistency in the results is the simultaneous activation of antagonistic muscles, i.e. GGP and HG, which can be seen for vowel [a]. This may be improved by adding more constraints to the tracking simulation not to consider solutions with antagonistic muscles.

As an alternative to EMG, Takano [34] used MRI to measure the length of extrinsic tongue muscles during vowel production. They found that the GG is the dominant muscle for the production of high-front vowels which is in agreement with our model estimations. According to this study, GGA, anterior part of HG, middle and posterior part of SG contribute for low-back vowels while GGP and GGM are relaxed.

For the jaw muscles, we found very small activity (1 percent in maximum). Since the jaw muscles are embedded for chewing and can generate very strong forces, we only need a small percent of activation to open or close the jaw in speech.

3.2. Articulatory postures and vocal tract geometry

Articulatory postures in Figure 5 show that the model succeeded to make the constriction in back, center and front part of the vocal tract. Furthermore, despite of the gaps between different articulators, an air-tight vocal tract was extracted. The area function was not used for acoustic simulations, but it is reported for comparison purposes against earlier studies in the literature.

3.3. Acoustic simulations

Figure 5 shows the vocal tract transfer functions $H(f)$ obtained by 3D FEM acoustic simulations. We observe the typical distribution of the first formants for each vowel sound, with frequency values (F_1 , F_2) in Hz of (760, 1512) for [a], (334, 998) for [u] and (266, 2245) for [i]. Note also that the formant amplitudes do not decay in frequency because of the definition of $H(f)$ in Eq. (2), which compensates somehow the glottal excitation. The typical 12dB decay/octave rate will

appear once a train of glottal pulses is introduced in the simulations so as to generate a vowel sound [30] (see a.wav, u.wav, and i.wav files for generated vowel sounds). Indeed, the last column of Figure 5 shows the spectrogram of the generated sounds when this train of glottal pulses is used in the FEM simulations. A pre-emphasis filter is applied to better visualize higher frequencies, as usually done in the literature. As expected, formant positions do not change over time, since the vocal tract was totally static during the simulations. However, small changes of formants over the time, which can be observed in the spectrograms, are produced by the pitch and amplitude variations of the introduced glottal pulses.

4. Conclusions and future work

We proposed a framework using articulatory data with a biomechanical model and FEM-based acoustic simulation to study speech production. A fully automatic method to couple biomechanics simulations with acoustic simulations was developed, and tested for cardinal vowel production. Although articulatory postures and acoustic simulations showed promising results, there still is uncertainty for the estimated activations of muscles. The framework can be used to study static sounds, i.e. vowels, and can be extended for diphthong sounds by solving the wave equation in a moving domain. For other categories of sounds, such as consonants and fricatives, a more advanced acoustic simulation is needed.

As future work, the tongue contour, from e.g. MRI or ultrasound, can be used instead of three sample points which may help to estimate more accurate muscle activation. Despite of EMA which is used directly, preprocessing is usually needed in other kind of articulatory data. Adding the lips to our adapted biomechanical model is another task to do in the future. We will also study diphthong sounds.

5. Acknowledgements

The authors would like to thank Peter Anderson for his assistance with the soft palate geometry. This research has been supported by EU-FET grant EUNISON 308874.

6. References

- [1] S. Maeda, "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modelling*, Springer, 1990, pp. 131–149.
- [2] P. Badin, G. Bailly, M. Raybaudi, and C. Segebarth, "A three-dimensional linear articulatory model based on MRI data," in *Third ESCA / COCOSDA International Workshop on Speech Synthesis*, 1998.
- [3] O. Engwall, "Combining MRI, EMA and EPG measurements in a three-dimensional tongue model," *Speech Commun.*, vol. 41, no. 2, pp. 303–329, 2003.
- [4] P. Birkholz, D. Jackel, and B. J. Kroger, "Construction And Control Of A Three-Dimensional Vocal Tract Model," in 2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings, vol. 1, pp. 1873–1876.
- [5] B. H. Story, "A parametric model of the vocal tract area function for vowel and consonant simulation," *J Acoust Soc Am*, vol. 117, no. 5, pp. 3231–3254, 2005.
- [6] Y. Payan and P. Perrier, "Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis," *Speech Commun.*, vol. 22, pp. 185–205, 1997.
- [7] J.-M. Gérard, R. Wilhelms-Tricarico, P. Perrier, and Y. Payan, "A 3D dynamical biomechanical tongue model to study speech motor control," arXiv Prepr. physics/0606148, 2006.
- [8] S. Buchaillard, P. Perrier, and Y. Payan, "A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning.," *J. Acoust. Soc. Am.*, vol. 126, no. 4, pp. 2033–2051, 2009.
- [9] X. Wu, J. Dang, and I. Stavness, "Iterative method to estimate muscle activation with a physiological articulatory model," *Acoust. Sci. Technol.*, vol. 35, no. 4, pp. 201–212, 2014.
- [10] Q. Fang, S. Fujita, X. Lu, and J. Dang, "A model-based investigation of activations of the tongue muscles in vowel production," *Acoust. Sci. Technol.*, vol. 30, no. 4, pp. 277–287, 2009.
- [11] I. Stavness, J. E. Lloyd, and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model.," *J. Biomech.*, vol. 45, no. 16, pp. 2841–8, Nov. 2012.
- [12] P. Anderson, N. M. Harandi, S. Moisiuk, I. Stavness, and S. Fels, "A Comprehensive 3D Biomechanically-Driven Vocal Tract Model Including Inverse Dynamics for Speech Research," in *INTERSPEECH*, 2015, pp. 2395–2399.
- [13] A. Toutios and S. Narayanan, "Articulatory synthesis of French connected speech from EMA data," in *INTERSPEECH*, 2013, pp. 2738–2742.
- [14] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J Acoust Soc Am*, vol. 100, no. 1, pp. 537–554, 1996.
- [15] D. Aalto, O. Aaltonen, R. P. Happonen, P. Jaasaari, A. Kivela, J. Kuortti, J. M. Luukinen, J. Malinen, T. Murtola, R. Parkkola, J. Saunavaara, T. Soukka, and M. Vainio, "Large scale data acquisition of simultaneous MRI and speech," *Appl. Acoust.*, vol. 83, pp. 64–75, 2014.
- [16] M. Zandipour, F. Guenther, J. Perkell, P. Perrier, Y. Payan, and P. Badin, "Vowel-vowel planning in acoustic and muscle space," in *Proceedings of "From Sound to Sense: 50+ years of discoveries in speech communication,"* 2004, pp. C103–C108.
- [17] J. M. Heinz and K. N. Stevens, "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech," in *Proceedings of the 5th International Congress on Acoustics*, 1965, p. A44.
- [18] P. Perrier, L. Boë, and R. Sock, "Vocal Tract Area Function Estimation From Midsagittal Dimensions With CT Scans and a Vocal Tract Cast Modeling the Transition With Two Sets of Coefficients," *J. Speech, Lang. Hear.*, vol. 35, no. 1, pp. 53–67, 1992.
- [19] P. Perrier, Y. Payan, S. Buchaillard, M. A. Nazari, and M. Chabanas, "Biomechanical models to study speech," *Faits de Langues*, vol. 37, pp. 155–171, 2011.
- [20] 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," *J Acoust Soc Am*, vol. 137, no. 2, pp. 832–843, 2015.
- [21] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, P. X., and A. Van Hirtum, "Effects of vocal tract geometry simplifications on the numerical simulation of vowels," 11th Pan-European Voice Conference (PEVOC). p. 177, 2015.
- [22] 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," *J Acoust Soc Am*, vol. 134, no. 4, pp. 2946–2954, 2013.
- [23] I. Stavness, C. A. Sánchez, J. Lloyd, A. Ho, J. Wang, S. Fels, and D. Huang, "Unified skinning of rigid and deformable models for anatomical simulations," in *SIGGRAPH Asia 2014 Technical Briefs*, p. 9.
- [24] 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," *Int. j. numer. method. biomed. eng.*, vol. 27, no. 3, pp. 367–390, 2011.
- [25] 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*, vol. 11, Springer, 2012, pp. 355–394.
- [26] A. Ben Youssef, "Control of talking heads by acoustic-to-articulatory inversion for language learning and rehabilitation," *GIPSA-lab / DPC, Doctorat de l'Université de Grenoble*. Grenoble: Grenoble University, 2011.
- [27] S. Dabbaghchian, M. Arnela, and O. Engwall, "Simplification of vocal tract shapes with different levels of detail," in 18th International Congress of Phonetic Science, 2015, pp. 1–5.
- [28] M. Arnela and O. Guasch, "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," *J. Acoust. Soc.*, vol. 133, no. 6, pp. 4197–4209, 2013.
- [29] P. Švancara and J. Horáček, "Numerical modelling of effect of tonsillectomy on production of Czech vowels," *Acta Acust. united with Acust.*, vol. 92, no. 5, pp. 681–688, 2006.
- [30] A. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Am.*, vol. 49, no. 2B, pp. 583–590, 1971.
- [31] T. Baer, P. J. Alfonso, and K. Honda, "Electromyography of the tongue muscles during vowels in/gpvp/environment," *Ann. Bull. RILP*, no. 22, pp. 7–19, 1988.
- [32] T. Gay, B. Lindblom, and J. Lubker, "Production of bite-block vowels: Acoustic equivalence by selective compensation," *J. Acoust. Soc. Am.*, vol. 69, no. 3, pp. 802–10, 1981.
- [33] K. Honda, "Organization of tongue articulation for vowels," *J. Phon.*, vol. 24, no. 1, pp. 39–52, 1996.
- [34] S. Takano and K. Honda, "An MRI analysis of the extrinsic tongue muscles during vowel production," *Speech Commun.*, vol. 49, no. 1, pp. 49–58, 2007.