Machine-Learning based model order reduction of a biomechanical model of the human tongue

Maxime Calka, Pascal Perrier, Jacques Ohayon, Christelle Grivot-Boichon, Michel Rochette, Yohan Payan

Machine-Learning based model order reduction of a biomechanical model of the human tongue

Maxime Calka[a,b,d], Pascal Perrier[b], Jacques Ohayon[a,c], Christelle Grivot-Boichon[d], Michel Rochette[d], Yohan Payan[a]

[a] *Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France*
[b] *Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France*
[c] *Savoie Mont-Blanc University, Polytech Annecy-Chambéry, 73376 Le Bourget du Lac, France*
[d] *ANSYS, F-69100 Villeurbanne, France*

Abstract

Background and Objectives: This paper presents the results of a Machine-Learning based Model Order Reduction (MOR) method applied to a complex 3D Finite Element (FE) biomechanical model of the human tongue, in order to create a Digital Twin Model (DTM) that enables real-time simulations. The DTM is designed for future inclusion in a computer assisted protocol for tongue surgery planning.

Methods: The proposed method uses an "a posteriori" MOR that allows, from a limited number of simulations with the FE model, to predict in real time mechanical responses of the human tongue to muscle activations.

Results: The MOR method is evaluated for simulations associated with separate single tongue muscle activations. It is shown to be able to account with a sub-millimetric spatial accuracy for the non-linear dynamical behavior of the tongue model observed in these simulations.

Email address: maxime.calka@univ-grenoble-alpes.fr (Maxime Calka)

Preprint submitted to Computer Methods and Programs in Biomedicine October 13, 2020

Conclusion: Further evaluations of the MOR method will include tongue movements induced by multiple muscle activations. At this stage our MOR method offers promising perspectives for the use of the tongue model in a clinical context to predict the impact of tongue surgery on tongue mobility. As a long term application, this DTM of the tongue could be used to predict the functional consequences of the surgery in terms of speech production and swallowing.
Keywords: Real-time simulation, Model Order Reduction, Digital Twins, Human tongue

1. Introduction

1.1. Medical context

Nowadays, tongue is the most common intraoral site for cancer [1]. In France, tongue cancer affects 4200 new patients each year [2] and all around the world it represents 30% to 50% of the oral cavity tumors [1, 3].

A common technique to treat patients suffering from tongue cancer is the exeresis of a part of the tongue [4]. This surgery can have severe consequences on tongue mobility and deformation capabilities, inducing impairments of mastication, deglutition and speech production which can reduce drastically the quality of life of patients [5]. Quantitative predictions of the functional consequences of this surgery is very complex for clinicians.

The present study is part of a long-term project aiming at developing a patient-specific "in silico" surgery planning system that should quantitatively predict the functional consequences of orofacial surgery. This will require:

- to automatically generate patient-specific 3D Finite Element (FE) tongue models [6].
- to quantitatively predict, within a time interval compatible with clinical constraints, the functional consequences of anatomical changes (e.g. tumor resection with flap reconstruction [7]) on swallowing and speech.

Achieving real-time simulations is the focus of our paper, and we propose for this to use a Model Order Reduction (MOR) method based on Machine-Learning techniques.

The tongue is a complex organ with incompressible tissues and nonlinear viscoelastic properties [8, 9]. Numerical simulations with an FE model of the human tongue, which accounts for the non-linear mechanical properties of tongue tissues and accurately implements tongue muscle anatomy, can take very long time (on an Intel(R) Xeon(R) with 16Gb and 8 logical cores about one hour to simulate a movement of some tens of milliseconds), which makes it difficult to use such a model in a clinical context [10, 11]. To study the functional outcome of the surgery in terms of speech production or swallowing, a key point is to be able to simulate tongue trajectories over time and not just to produce the final tongue shape resulting from muscles contractions. Hence a transient FE analysis, which solves movement equations, is required. In this context a difficulty is that tongue can move quite rapidly in speech production (10 to 20 cm/s), which increases the impact of visco-elastic properties on movement. To account for this phenomenon, the challenge of MOR techniques is to capture the non-linear behavior of the tongue.

1.2. Related Works

MOR methods have recently received a growing interest to challenge the real-time simulation problem in computer-aided surgery [12]. These methods allow to obtain real-time simulations by reducing the computational complexity without simplifying the physics of the model.

Projection-based and collocation-based MOR methods are the most popular ones [13, 14, 15, 16]. Projection-based MOR methods are divided in two categories: (1) *a posteriori* methods, such as the Proper Orthogonal Decomposition, which create a Reduced Order Model (ROM) from a large set of simulations called "Snapshots" [13] and require a computationally intensive *offline* phase; (2) *a priori* methods, such as the Proper Generalized Decomposition, which reduce the model during the problem solving process itself [14, 17]. A method of the former type was applied to computational medicine by Niroomandi et al. [15] in the case of non-linear quasi-static and large deformation problems to simulate the palpation of the human cornea.

In [16] the ROM was created with a collocation-based MOR method called Space Subspace Learning [18]. The authors have developed a Digital Twin Model (DTM) of the liver by considering the large displacement approach in linear elasticity and quasi-static way.

1.3. Overview

Unlike previous studies, a strong constraint in physical modeling of tongue in speech production and swallowing is the necessity to solve movement equations over time. To do so we propose to rely on an *a posteriori* machine-learning-based MOR (ML-based MOR) method using a recurrent neural network. We tested the capacity of the ML-based MOR method to account for

the dynamical characteristics of the tongue, by designing and evaluating two reduced models of the tongue in the aim to account for the movements of the tongue in response to the separate activations of two important tongue muscles, which are associated to two different kinds of tongue displacements. Below, the MOR method and the biomechanical tongue model are described. Then, the FE simulations, which resulted from these two separate muscle activations and served in the learning phase of the MOR are presented, and the capacity of the ROM to estimate tongue displacements in response to theses single muscle activations is evaluated.

2. Materials & Methods

2.1. ML-based MOR

In sum, we expect the ROM to functionally accounts for the dynamical behavior of the biomechanical tongue model over time. Tongue deformation over time is induced in the biomechanical model by the time variations of m muscle commands $\{g_k(t), k \in [1, m]\}$, called inputs. This tongue deformation is described with n time varying spatial coordinates of the nodes located on the surface of the tongue model, $\{p_i(t), i \in [1, n]\}$, called outputs. Input and output variables are sampled at n_t regular time steps during the course of the movement. Thus, the mechanical response of the biomechanical tongue model to muscle commands is described by two matrices, the input matrix G_{m,n_t} called "excitation" and the output matrix P_{n,n_t}. In the FE formulation this input-output relation is computed with a full-order transient solver. In the MOR this relation has to be learned from a limited number of sets $\{\hat{P}_{n,n_t}, \hat{G}_{m,n_t}\}$ (called scenarios), in order to build a ROM that accounts for

the observed scenarios and generalizes the relation to unknown situations. We used the MOR technique developed by ANSYS® called "Dynamic ROM Builder" (DRB), which is accessible in the ANSYS Twin Builder product [19]. The DRB algorithm is being patented[1]. The modeling process is described in Figure 1 and consists of two steps: (i) an offline phase involving, first, a reduction of dimensionality of the output vector using a Singular Value Decomposition (SVD) that generates new variables $\hat{X}_{r,n_t}$ from the original variables $\hat{P}_{n,n_t}$ and, second, the ROM building using a learning method that optimizes the structure of the ROM from the set of $\hat{X}_{r,n_t}$ variables ; (ii) an online phase in which mechanical responses can be generated from new values of the muscle commands G_{m,n_t}, first using the ROM which generates variables $X_{ROM_{r,n_t}}$ in the space resulting from the dimensionality reduction induced by the SVD carried out on the original data, and second by transforming the outputs of the ROM into estimations of the surfacic tongue nodes coordinates $P_{ROM_{n,n_t}}$ via an inverse transform of the SVD.

2.1.1. Learning phase: ROM building

The DRB algorithm models the dynamical behavior of the biomechanical tongue model, as described by variable X, with two equations:

$$\dot{X}(t) = f(X(t), G(t)) \tag{1}$$

$$X(0) = X_0 \tag{2}$$

[1]This application can be referenced as US Patent Application No. 16,527,387, entitled SYSTEMS AND METHODS FOR BUILDING DYNAMIC REDUCED ORDER PHYSICAL MODELS, filed July 31, 2019.

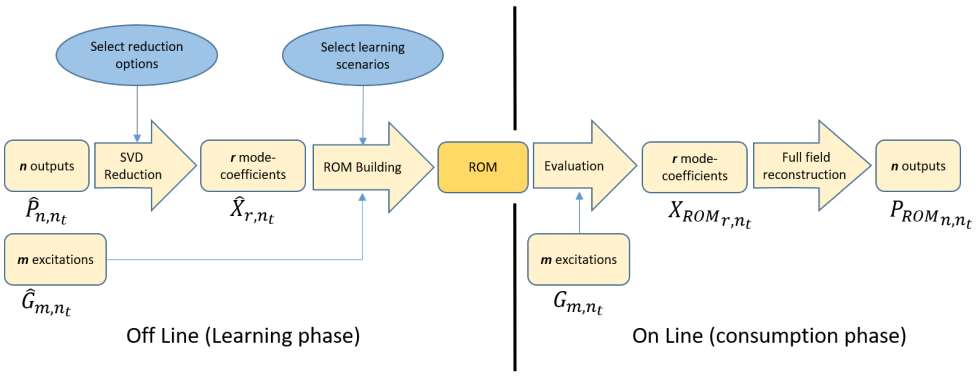


Figure 1: Overview of the different steps performed during the simulation of the Dynamic ROM Builder process. Variables G (dimension m, n_t) are the excitations that generate tongue deformations ($\hat{G}_{m,n_t}$ are the excitations associated with the learning scenarios; G_{m,n_t} are the excitations associated with the simulations with the ROM). Variables P (dimension n, n_t) describe the variation of the coordinates of the n nodes on the surface of the tongue model over the n_t time steps of the simulations ($\hat{P}_{n,n_t}$ are the data included in the learning scenarios; $P_{ROM_{n,n_t}}$ are the coordinates of the nodes resulting from the simulations with the ROM). Variables X (dimension r, n_t, with r smaller than n) are the mode coefficients, which are in the space resulting from the dimensionality reduction applied to the space of the surfacic tongue nodes thanks to the SVD (see equation (6)) ($\hat{X}_{r,n_t}$ results from the SVD applied to $\hat{P}_{n,n_t}$; $X_{ROM_{r,n_t}}$ is the output of the ROM, which is transformed into $P_{ROM_{n,n_t}}$ via the inverse transform of the SVD as shown in equation 7).

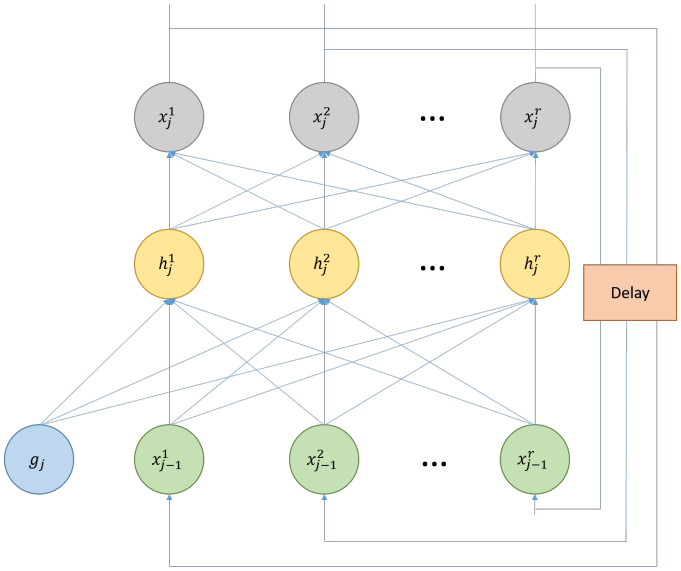
- $X(t)$ is the output vector (dimension r) at time t of the ROM
- $\dot{X}(t)$ is the first derivative of $X(t)$
- $G(t)$ is the input vector (dimension m) at time t
- f is a non-linear function which has to be learned from the set of variables $\{\hat{X}_{r,n_t}, \hat{G}_{m,n_t}\}$ corresponding to the learning scenarios.

The learning process aims at finding the non-linear function f that minimizes the average quadratic error E (equation 3), computed over the full set of learning scenarios, between the variables $\hat{X}_{r,n_t}$ computed with the full-order transient FE model and the outputs X_{r,n_t} of the ROM predicted with equations (1) and (2) for the input vectors $\hat{G}_{m,n_t}$:

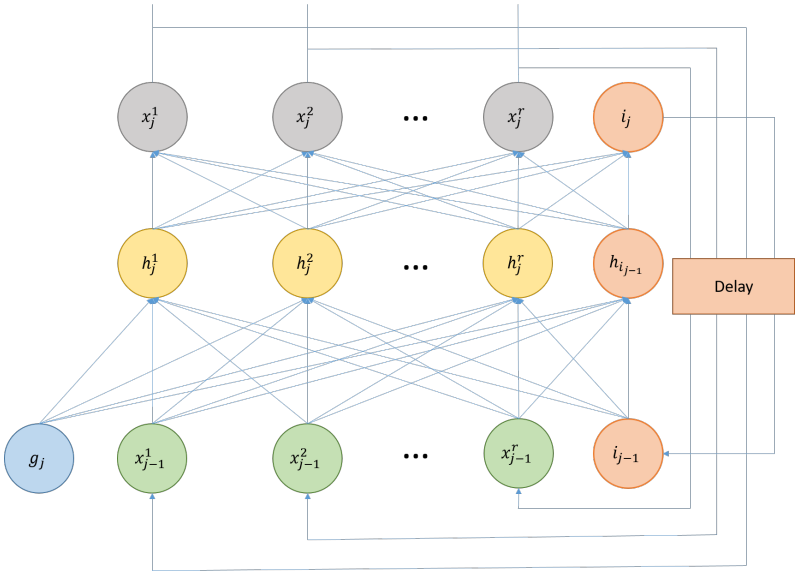
$$E = \overline{\left(\frac{1}{r} \sum_{l=1}^{n} \left(\frac{1}{n_t} \sum_{j=1}^{n_t} (X_{l,j} - \hat{X}_{l,j})^2 \right) \right)} \tag{3}$$

Function f is a quadratic function. It is implemented as a 3-layer recurrent neural network with the same number of variables in the hidden and the output layers (see Figure 2a).

The activation function used in the hidden and output layers is a sigmoid. Conventional gradient descent optimization methods [20] are used in the optimization process of f, which stops when error E (equation 3) becomes smaller than a predefined user-dependent threshold ε. If this threshold cannot be reached, a free variable $i_j, j \in [1, n_t]$, is added in the neural network implementation of f. A free variable can be considered as a memory cell in the neural network, which is external to the set of input and output variables, and which value is adapted along with the other parameters of the network at each time step. Adding a free variable in the network is done by adding an



(a) Original neural network



(b) Addition of a free variable

Figure 2: Scheme of the recurrent neural network implementation (at time step j). Panel a: Usual 3-layer recurrent network representation; Panel b: 3-layer recurrent network including free-variables to account for high-frequency dynamical properties.

output cell i_j and a corresponding input/hidden cell in the recurrent neural network (see Figure 2b).

Each new free variable is initialized as the time-varying error averaged on all the network outputs: $i_j = \frac{1}{r}\sum_{l=1}^{r}(X_{l,j} - \hat{X}_{l,j})$ with j varying from 1 to n_t. Then, the optimization process starts over. If again the required threshold error cannot be reached, another free variable is added, using the same procedure, and the optimization process starts over. This procedure is repeated as many times as necessary until the (minimum) required error threshold ε is reached. Assuming k iterative steps, in which k free variables are added, the optimal ROM models the dynamical behavior of the FE tongue model according to equation (4),

$$\begin{pmatrix} \dot{X}_j \\ \dot{I}_j \end{pmatrix} = f\left(\begin{pmatrix} X_j \\ I_j \end{pmatrix}, G_j \right), j \in [1, n_t], \tag{4}$$

in which I_j and $\dot{I}_j$ are k dimensional vectors corresponding to the free variables that were generated along the iterative optimization process and their first time derivatives. The inclusion of the free variables is the innovative part of the DRB method. It enables us to obtain a better approximation of dynamical behavior of the biomechanical tongue model by accounting for complex non-linearities and higher order time-dependency characterizing this behavior, without increasing the depth (i.e. the number of layers) of the recurrent neural network, which avoids "vanishing gradients" problems [21, 22].

2.1.2. Output vector: Reduction of dimensionality

Ultimately, the outputs of the ROM should enable to generate at each time-step an accurate approximation of the vector P of the coordinates of the surfacic nodes of the biomechanical tongue model. Hence, at a first glance it would be natural to design the ROM directly from the $\hat{P}_{n,n_t}$ matrices of the learning scenarios, and, in turn, to have the ROM generate directly estimations of P_{n,n_t} coordinates in the consumption phase. However, given the high mesh density required for accurate simulations, the dimensionality n of vector P is very high. Using this vector as output of the recurrent network of Figure 2 would induce a considerable computational complexity for the learning phase. To reduce this complexity of the output space, the DRB method uses Singular Value Decomposition (SVD). SVD was chosen instead of recent and statistically more powerful techniques, such as autoencoders, because it makes it possible to keep the physical components that are the most influential on tongue movements, such as inertia, incompressibility, and the fundamental law of dynamics, and to eliminate components related to computational inaccuracy without reliable physical foundations. SVD enables us to reduce the dimensionality of the output matrix by first decomposing the matrix $\hat{P}_{n,n_t}$ of the coordinates of the n surfacic nodes at the n_t time steps of the learning scenarios as follows:

$$\hat{P}_{n,n_t} = U_{n,n} \cdot \Sigma_{n,n_t} \cdot V^{\intercal}_{n_t,n_t} \tag{5}$$

where U and $V^{\intercal}$ are unitary matrices corresponding to the left and right singular vectors of $\hat{P}_{n,n_t}$ and Σ_{n,n_t} is a diagonal matrix which terms are the singular values of matrix $\hat{P}_{n,n_t}$, ordered in descending magnitude from the

first to the last line. This decomposition allows us to do an approximation of $\hat{P}_{n,n_t}$ by setting to zero the singular values that are smaller than a given threshold. Thus the dimensionality of Σ_{n,n_t} is reduced to (r, r) and matrix $\hat{P}_{n,n_t}$ is approximated by $\tilde{P}_{n,n_t}$ as follows:

$$\tilde{P}_{n,n_t} = U_{n,r} \cdot \Sigma_{r,r} \cdot V^{\intercal}_{r,n_t} \tag{6}$$

Hence, the coordinates of the n surfacic nodes are approximated with enough accuracy on the basis of the first r left singular vectors, called modes. To these modes are attached at each time-step j, $j \in [1, n_t]$, r mode-coefficients $\hat{X}_{r,j}$ that are computed, consistent with equation (6), with equation (7):

$$\hat{X}_{r,n_t} = \Sigma_{r,r} \cdot V^{\intercal}_{r,n_t}. \tag{7}$$

In the learning phase of the ROM, the recurrent neural network (Figure 2) is optimized in order for its outputs to satisfactorily approximate the matrix $\hat{X}_{r,n_t}$ over the whole set of scenarios.

Once the ROM is learned, for the simulations with the ROM, matrix P_{n,n_t} of the n surfacic nodes of the biomechanical tongue model is estimated from the output matrix X_{r,n_t} of the ROM, in agreement with equation (6), by multiplying X_{r,n_t} with the matrix $U_{n,r}$ of the r first left singular vectors of $\hat{P}_{n,n_t}$:

$$P_{ROM_{n,n_t}} = U_{n,r} \cdot X_{r,n_t} \tag{8}$$

Importantly, SVD provides a linear account of the spatial relation between surfacic nodes, whereas tongue strain in response to stress is known to obey non-linear mechanical laws. Despite this apparent contradiction,

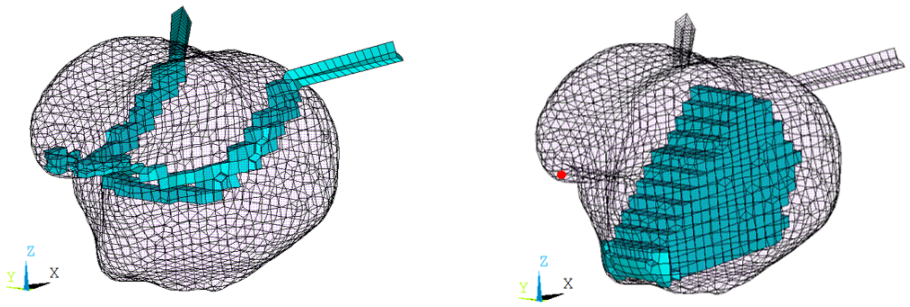
SVD has for our modeling work a crucial feature: the physical phenomena responsible for the tongue movement characteristics of largest magnitudes, namely the mass, the stiffness and the damping factor, are represented by the largest singular values. They also correspond to low-frequency modes of the mechanical system. Thus, SVD essentially keeps low frequency modes. Consequently, the prediction error of SVD, i.e. the difference between the actual time-varying positions of the surfacic nodes ($\hat{P}_{n,n_t}$) and their lower dimensional account ($\tilde{P}_{n,n_t}$) after SVD, mainly includes high frequency components. Importantly, these high-frequency components are the consequence of different phenomena, of which only a part actually reflects the true complexity of the physical properties of the tongue, which the SVD cannot account for faithfully because of its linearity properties. Another part is due to high frequency computational noise intrinsically associated with FE solvers, which it is in fact interesting not to integrate in the modeling since they do not correspond to real characteristics of the tongue. By selecting the low-frequency modes, SVD removes high frequency noise that has no physical meaning, but also those resulting from the physics. The DRB approach uses free variables to put back into the model physically meaningful high-frequency components. Indeed, it is believed that consistency in dynamical behavior of the biomechanical model across simulations makes it likely that the high frequency components included in the added free variables account mainly for real physical phenomena.

2.1.3. 3D biomechanical model of the human tongue

The tongue model used for the simulations is described in [23]. It is based on an FE mesh with 7763 nodes and 8780 hexahedral elements. The

constitutive law used to model the elastic properties of the tongue tissue is a Mooney-Rivlin material with two parameters C_{10} and C_{20} (respectively equal to 192 Pa and 90 Pa). Tongue viscosity is approximated with a Rayleigh model (Rayleigh coefficients: $\alpha = 20$ and $\beta = 0.0$). To model the quasi-incompressibility of the tissue, the Poisson ratio is fixed to $v = 0.4999$. No-displacement "boundaries conditions" are defined on the nodes in contact with the jaw and on the lowest boundary of the mouth floor.

Figure 3 shows the mesh of the tongue model and highlights in blue the two muscles which will be independently activated in the numerical simulations used in the learning phase, namely the styloglossus (SG) and the genioglossus posterior (GG-P).



(a) Styloglossus (b) Genioglossus posterior

Figure 3: FE mesh of the tongue model used for the simulations with the two activated muscles highlighted in blue. Each muscle activation is modeled as a transversely isotropic material with an activation along the main direction of the fibers. The red point on the tongue tip, located in the mid-sagittal plane of the tongue, is used below to illustrate the accuracy of the predictions with the ROM.

2.2. Simulation data

The learning phase is based on a set of scenarios. Each scenario consists of two sets of data called "excitation" (inputs) and "output". The excitation $\hat{G}_j, j \in [1, n_t]$ is a time-varying activation of one of the two considered muscles and the output $\hat{X}_j, j \in [1, n_t]$ corresponds to the r mode-coefficients computed from the coordinates of the surfacic nodes (1861 nodes) according to equation (7).

2.2.1. Excitation

In this study we have built two different ROMs of the biomechanical tongue model corresponding to two quite different kinds of deformations. The ROM was learned from simulations of tongue movements in response to the activation of an intrinsic muscle located in the center of the tongue, the Genioglossus Posterior (GG-P), which is responsible for protrusion and elevation of the tongue [10]. The second ROM was learned from simulations of tongue movements in response to the activation of an extrinsic muscle, the Styloglossus (SG), which raises and retracts the tongue [10]. Thus two sets of excitations are studied: (1) activations of the GG-P muscle alone; (2) activations of the SG muscle alone. In both cases, muscle activation patterns consist of a linearly increasing phase followed by a stabilization phase (Figure 4). This approach does not aim at building a unique ROM of the tongue, which could account for every kind of tongue deformation associated with any pattern of muscle activations (such an objective would require extensive coverage of the motor command space), but, more modestly, to assess the capacity of the DRB method to account for different complex non-linear time-deformations of the tongue along different directions. This is an essen-

tial prerequisite for any further effort to build a unique and exhaustive ROM of the tongue.

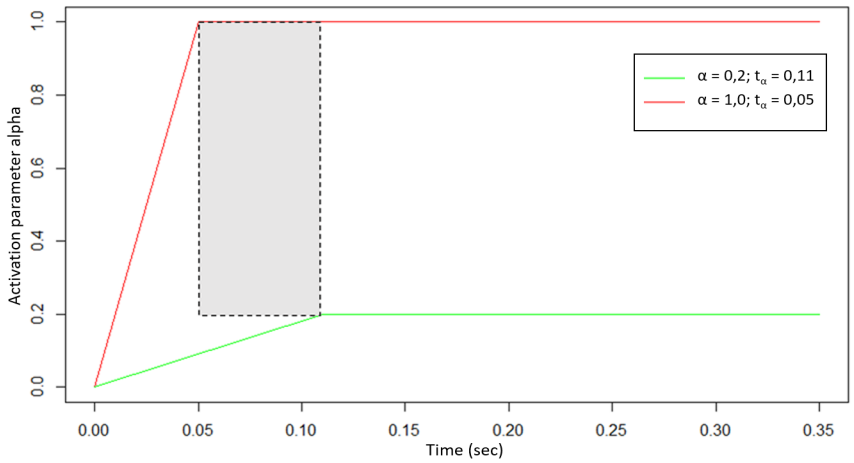


Figure 4: Range of variation of all possible excitation patterns used in the scenarios, either with GG-P or with SG . In red, the pattern corresponding to the maximum stress in the stabilization phase with a minimum duration of the increasing phase of muscle activation. In green the pattern corresponding to the minimum stress in the stabilization phase with a maximum duration of the increasing phase of the activation. The rectangle in grey corresponds to the whole range of possible parameters values of the simulations.

In the FE model muscle activation is directly defined as a stress that increases from zero to the value σ (expressed in Pa) reached in the stabilization phase. σ is specified in reference to a maximum value σ_{max} via an activation parameter α in the interval $[0; 1]$ such that $\sigma = \alpha \times \sigma_{max}$. In our scenarios α varies in the range $[0.2; 1.0]$. All the simulations have a total duration t_{total} of $0.35\,s$ with a duration of the initial linearly increasing phase t_α in the range $[0.05\,s, 0.11\,s]$. These durations have been chosen because they correspond to the generation of realistic tongue movements in speech production with

the biomechanical model.

Figure 5 illustrates how the non-linear dynamics of tongue tissue shapes the kinematics of the tongue, with the displacement along the 2 axes of the mid-sagittal plane of a point located on the tip of the tongue in the mid-sagittal plane (red dot on Figure 3) during a GG-P activation. The model being symmetrical nodes located in the mid-sagittal plane do not move along the y direction orthogonal to the mid-sagittal plane.

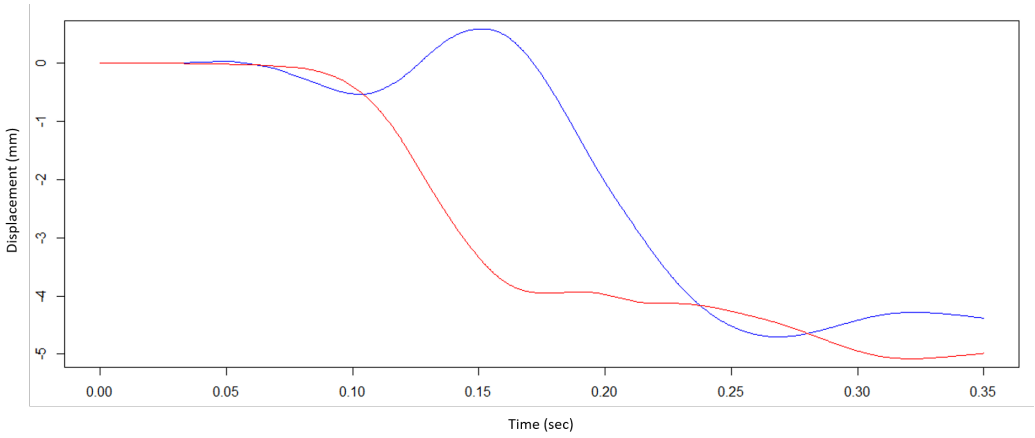


Figure 5: Tongue displacement of a node (represented Figure 3 in red) after activation of the GG-P. Red: Front back horizontal direction x; Blue: Vertical direction z. $\alpha = 0.6, t_\alpha = 0.09\,s$

2.2.2. Output

The 3D motion of the 1861 surfacic nodes is used to evaluate the performance of the ROM.

2.3. Learning scenarios

Two sets of 20 simulations were conducted, one for the GG-P and one for the SG, in order to set up the learning scenarios. These simulations were

performed with excitation data whose parameters α and t_α had steps of 0.2 and 0.02 s respectively within the ranges of variation given above, forming a grid represented on Figure 6.

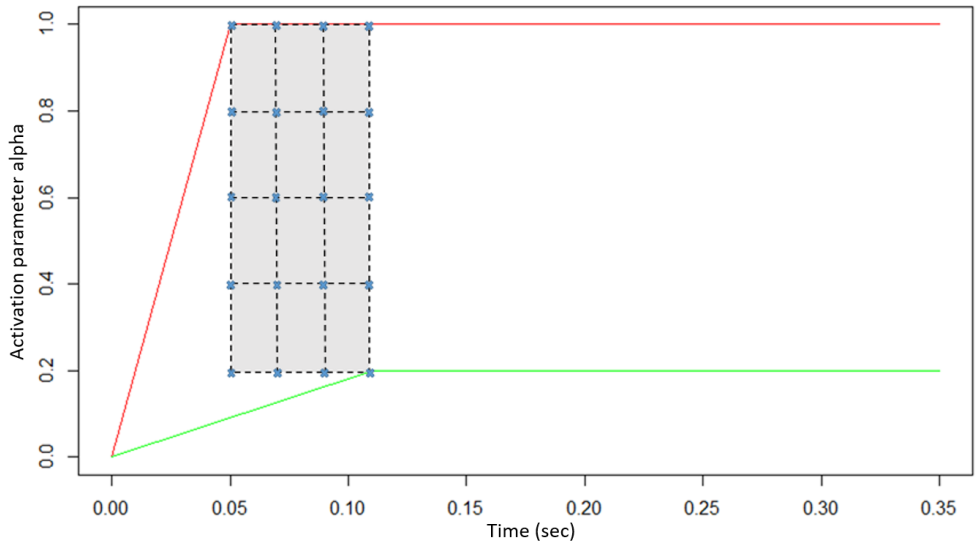


Figure 6: Set of learning scenarios for each activation cases (SG, GG-P)

For the learning of the two ROMs we set minimum error threshold ε to obtain an accuracy of less than 1/10mm (see section 2.1.1), which enabled us to have an average root mean square error on the evaluation scenarios (see below) in the order of a few tenths of millimeters.

2.4. Evaluation scenarios

Two sets of 20 simulations were used for the evaluation scenarios. Parameters α and t_α were randomly distributed inside four subparts of the grid of Figure 7 to cover a sufficiently large range of possibilities without using any set of parameters already used in simulations that served for learning.

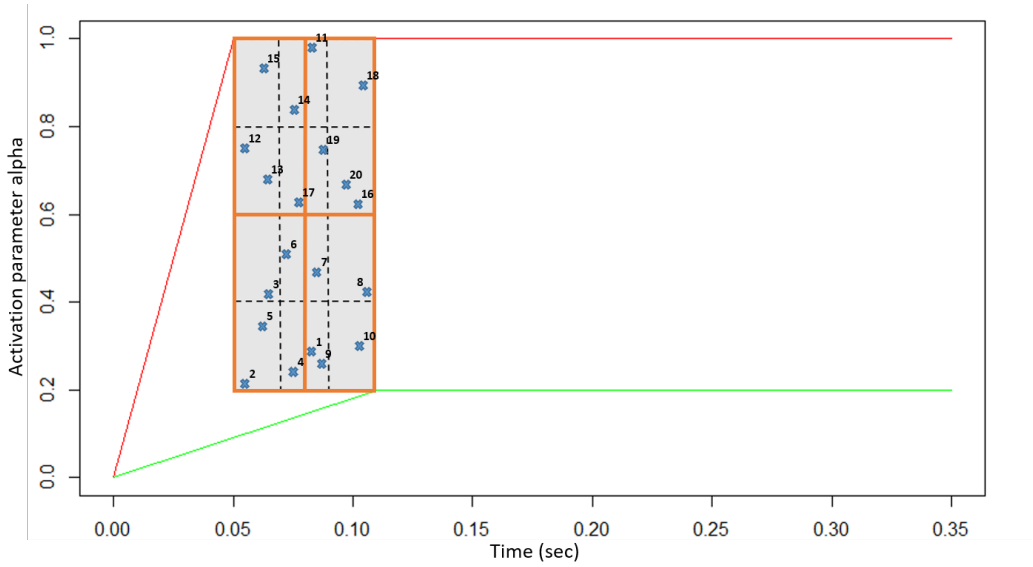


Figure 7: Set of excitations used in the evaluation scenarios for a single muscle

2.5. Metric

To quantitatively evaluate the ROM accuracy, the metric used is the average root mean square error (average RMS error) between the coordinates of the surfacic nodes as approximated with the ROM and the coordinates computed with the full-order transient FE model. It is computed with a formula similar to the one of equation 3, in which $\hat{X}_{n,n_t}$ and X_{n,n_t} are replaced by the matrices of the approximated and ground-truth coordinates (P_{ROM} and P) of the nodes. An average RMS error of less than a few tenths of a millimeter is considered to reveal a satisfactory quality.

3. Results & Discussion

3.1. Results

Figure 8 shows the original and the approximated x and z displacements of a node on the tongue tip (red dot on Figure 3) associated with the ROM learned from the activations of the GG-P alone (Panel a) and with the model learned from the activations of the SG alone (Panel b). In both cases, only one free variable was required for the optimal ROM. We see, on these examples, that the DRB method provides a good approximation of the ground-truth deformations of the biomechanical tongue model over time in response to either the GG-P or the SG muscle.

The average and standard deviation across surfacic nodes of the RMS error computed over the whole movement are given in Figure 9 for each of the 20 scenarios separately. Depending on the scenario the RMS error varies between 0.038 mm and 0.074 mm for the GG-P activation, and between 0.084 mm and 0.146 mm for the SG activation. Standard deviation is between 0.011 mm and 0.018 mm for the GG-P and between 0.028 mm and 0.062 mm for the SG. This is in the order of magnitude of the accuracy reached by the most sophisticated tongue movement tracking systems such as Electro Magnetic Articulometer (EMA) [24].

Figure 10 illustrates for 4 evaluation scenarios associated with 4 increasing levels of each muscle activation the spatial distribution over the surfacic nodes of the prediction error, computed as the module of vector $(P - P_{ROM})$. We observe that the prediction error is quite evenly distributed and is in general low, except in the posterior velar region where the external branches of the SG, arising from the styloid process, enter the body of the tongue. The

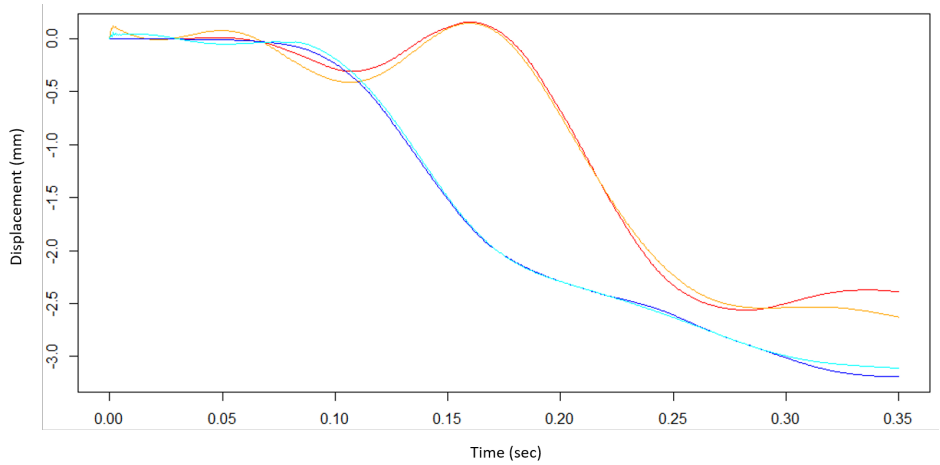
external branches of the SG are not represented in Figure 10a in order to focus on the error map on the tongue body, which determines vocal tract geometry. This strong and localized prediction error deserves further investigation.

We assessed the stability of mechanical equilibrium predicted with the ROM in the stabilization phase by extending the duration of this phase in new simulations. Figure 11 shows the vertical displacement of the tongue tip node generated with the ROM (learned with a $t_{total} = 0.35\,s$) for a total duration t_{total} of one second, with a SG activation of $\alpha = 0.24, t_\alpha = 0.076\,s$. The simulated movement is stable with an RMS error averaged on the surfacic nodes of 0.29 mm and a standard deviation of 0.19 mm.

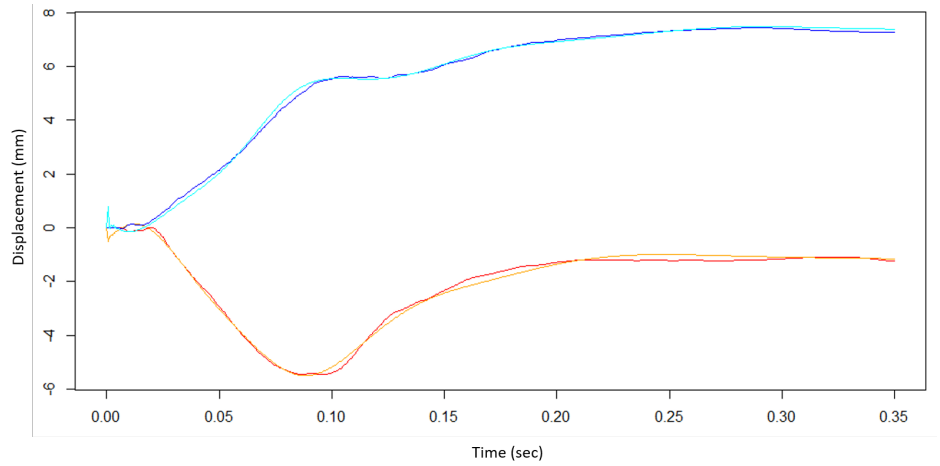
3.2. Discussion

Using the DRB method we have designed two different ROM in order to account for the deformations of the tongue in response to the separate activations of the GG-P and the SG. The results are encouraging as concerns the capacity of the DRB method to account for the dynamical behavior of tongue tissues. Each ROM generates in real time tongue movements that are close to those generated with the original biomechanical model, with a sub-millimetric average RMS error. Slight differences are observed in the approximation quality between the ROM based of GG-P activations and the one based on SG activations. Figure 8 provides a possible explanation: the trajectory generated with the biomechanical model is more noisy for the activation of the SG, probably because of some numerical inaccuracies.

Figure 8 suggests that in the considered scenarios the node trajectories are not so complex, and are similar to the indicial response of an under-damped second order system. This is consistent with the fact that only one

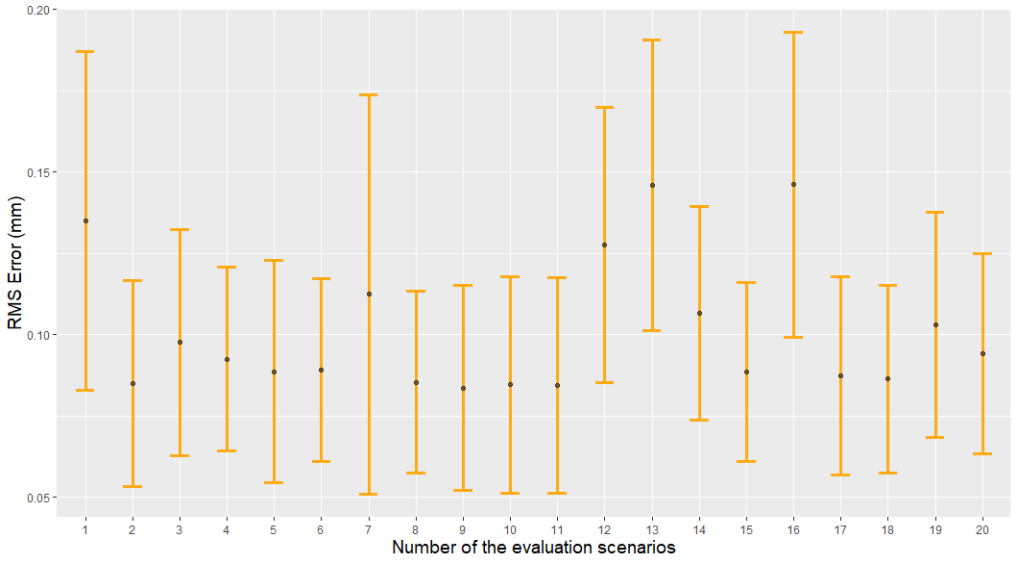


(a) Genioglossus posterior

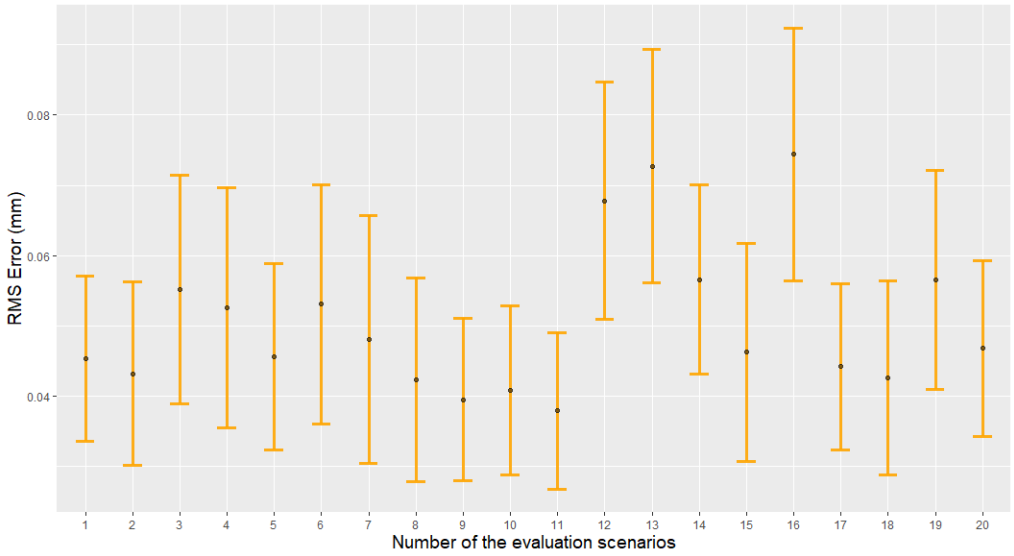


(b) Styloglossus

Figure 8: Examples of displacements of a node on the tip of the tongue (red point represented in Figure 3) in response to an activation of the GG-P alone (Panel a) and an activation of the SG alone (Panel b). Red curves: displacement resulting from the simulations with the biomechanical model along the vertical direction z. Orange curves: displacement computed with the ROM along z. Blue curves: displacement resulting from the simulations with the biomechanical model along the front-back horizontal direction x. Cyan curves: displacement computed with the ROM along x. In both figure, $\alpha = 0.29, t_\alpha = 0.081\, s$

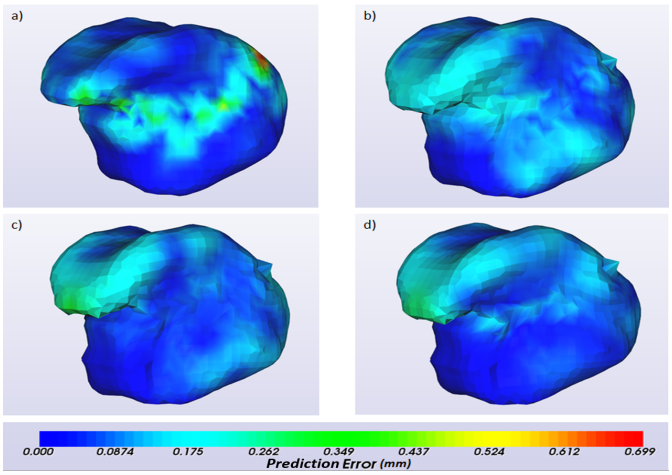


(a) Activations of the styloglossus alone

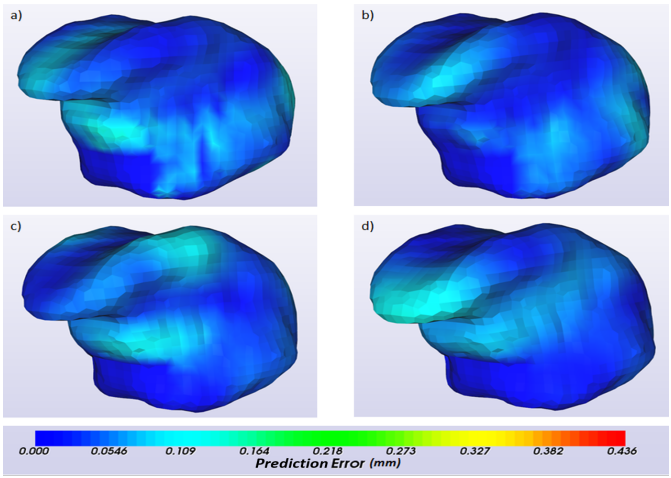


(b) Activations of the genioglossus posterior alone

Figure 9: Average (black dot) and standard deviation (orange segment around the dot) across nodes of the RMS error computed over the whole movement for each of the 20 scenarios.



(a) Prediction error for styloglossus activations



(b) Prediction error for genioglossus posterior activations

Figure 10: Distribution (Heat Map) over the nodes on the tongue surface of the prediction error (in mm, see text) for the activation of the SG alone (Top panel) and the GG-P alone (Bottom panel), corresponding to 4 increasing levels of activation (from (a) to (d)). The heat map representing this distribution is superimposed on the tongue shape achieved at the corresponding time of the movement. The activations are defined such as: a) ($\alpha = 0.29$, $t_\alpha = 0.081\,s$), b) ($\alpha = 0.44$, $t_\alpha = 0.085\,s$), c) ($\alpha = 0.69$, $t_\alpha = 0.065\,s$), d) ($\alpha = 0.88$, $t_\alpha = 0.104\,s$).

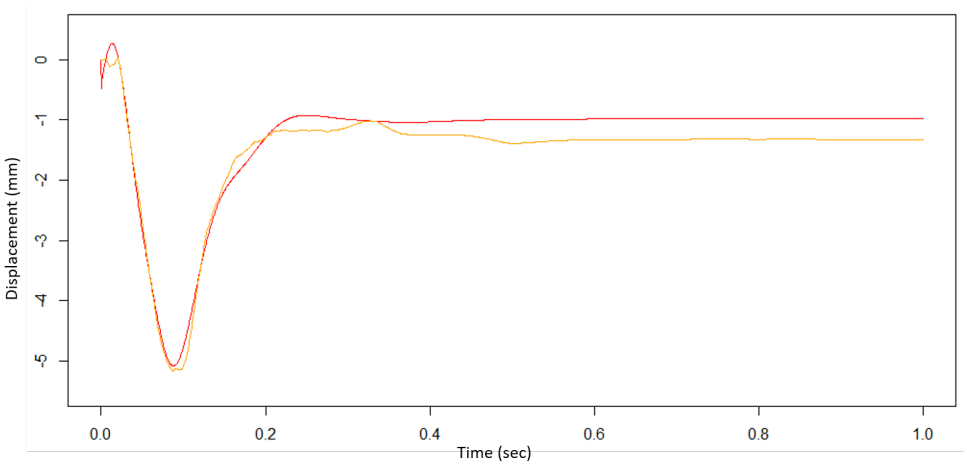


Figure 11: ROM estimation of the displacement a node on the tip of the tongue (red dot on Figure 3) over a total duration $t_{total} = 1\,s$ along the z axis for a SG activation with $\alpha = 0.24, t_\alpha = 0.076\,s$

free variable was required for the optimal ROMs and it could explain the good approximation that we have obtained at this stage for both ROMs. Further work will involve more complex muscle activation scenarios in which mechanical non-linearities will have stronger consequences on tongue movements. We will consider cases in which several muscles are activated at the same time, with different timings, as well as situations involving contacts between tongue and vocal tract boundaries.

Importantly, Figure 9 & 10 show that the final configurations are predicted accurately. This makes us confident in the capacity of the ROM to reliably assess, in the context of clinical applications, the range of speech articulations that a patient will be able to produce in post-surgical conditions. In addition, the Figure 11 shows that the ROM is able to provide accurate and stable predictions beyond durations used in learning. This sug-

gests that the MOR method learns the dynamics of the tongue model and not only statistical relations between inputs and outputs as given in the learning scenarios.

4. Conclusion

Two distinct ROMs have been constructed on two kinds of muscle activations (SG alone and GG-P alone) with the DRB method, which relies on a 3-layer recurrent neural network and the original use of free variables. This method doesn't increase the deepness of the neural network, which has the advantage of not requiring deep-learning methods based on large data sets with possible vanishing gradient problems. Our first results show that the designed ROM predicts tongue movements in response to single muscle activations in real time with a sub-millimetric average accuracy.

Further evaluations are required in situations where stronger non-linearities are involved and more complex muscle activation patterns are used. The ultimate goal is to have a unique ROM accounting for the global dynamics of our biomechanical tongue model in response to any pattern of muscle activations. This is a basic requirement in order to use this ROM method with models of resected and reconstructed tongues of patients included in clinical protocols.

5. Acknowledgments

Sincere thanks to Mohamed Masmoudi, co-inventor of the DRB algorithm, for fruitful discussions and helpful advices. This work was partly supported by the French ANR within the Investissements d'Avenir program un-

der references ANR-11-LABX-0004 (Labex CAMI) and by MIAI@Grenoble Alpes (ANR-19-P3IA-0003)

References

[1] S. Moore, N. Johnson, A. Pierce, D. Wilson, The epidemiology of tongue cancer: a review of global incidence, Oral Diseases 6 (2000) 75–84.

[2] K. Jéhannin-Ligier, E. Dantony, N. Bossard, F. Molinié, G. Defossez, L. Daubisse-Marliac, P. Delafosse, L. Remontet, Z. Uhry, Projection de lincidence et de la mortalité par cancer en france métropolitaine en 2017, Rapport technique. Saint-Maurice: Santé publique France (2017) 1–80.

[3] A. El Bousaadani, M. Abou-Elfadl, R. Abada, S. Rouadi, M. Mahtar, M. Roubal, M. Essaadi, F. Kadiri, Cancer de la langue: épidémiologie et prise en charge, Journal Africain du Cancer/African Journal of Cancer (2015) 1–5.

[4] J.-M. Prades, T. Schmitt, A. Timoshenko, Cancers de la langue, EMC-Oto-rhino-laryngologie 1 (2004) 35–55.

[5] Z. hui Yang, W. liang Chen, H. zhang Huang, C. bin Pan, J. song Li, Quality of life of patients with tongue cancer 1 year after surgery, Journal of Oral and Maxillofacial Surgery 68 (2010) 2164 – 2168.

[6] A. Bijar, P.-Y. Rohan, P. Perrier, Y. Payan, Atlas-based automatic generation of subject-specific finite element tongue meshes, Annals of Biomedical Engineering 44 (2016) 16–34.

[7] K. Kappert, M. van Alphen, S. van Dijk, L. Smeele, A. Balm, F. van der Heijden, An interactive surgical simulation tool to assess the consequences of a partial glossectomy on a biomechanical model of the tongue, Computer Methods in Biomechanics and Biomedical Engineering 22 (2019) 827–839. PMID: 30963800.

[8] A.-A. K. Yousefi, M. A. Nazari, P. Perrier, M. S. Panahi, Y. Payan, A visco-hyperelastic constitutive model and its application in bovine tongue tissue, Journal of Biomechanics 71 (2018) 190 – 198.

[9] J.-M. Gérard, J. Ohayon, V. Luboz, P. Perrier, Y. Payan, Non-linear elastic properties of the lingual and facial tissues assessed by indentation technique: application to the biomechanics of speech production, Medical Engineering & Physics 27 (2005) 884–892.

[10] S. Buchaillard, P. Perrier, 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 126 (2009) 2033–2051.

[11] M. A. Nazari, P. Perrier, Y. Payan, The distributed lambda (λ) model (dlm): A 3-d, finite-element muscle model based on feldman's λ; model; assessment of orofacial gestures, Journal of Speech, Language, and Hearing Research 56 (2013) 1909–1923.

[12] E. Cueto, F. Chinesta, Real time simulation for computational surgery: a review, Advanced Modeling and Simulation in Engineering Sciences 1 (2014) 11.

[13] A. Chatterjee, An introduction to the proper orthogonal decomposition, Current Science 78 (2000) 808–817.

[14] F. Chinesta, R. Keunings, A. Leygue, The proper generalized decomposition for advanced numerical simulations: a primer, Springer Science & Business Media, 2013.

[15] S. Niroomandi, I. Alfaro, E. Cueto, F. Chinesta, Accounting for large deformations in real-time simulations of soft tissues based on reduced-order models, Computer Methods and Programs in Biomedicine 105 (2012) 1 – 12.

[16] N. Lauzeral, D. Borzacchiello, M. Kugler, D. George, Y. Rmond, A. Hostettler, F. Chinesta, A model order reduction approach to create patient-specific mechanical models of human liver in computational medicine applications, Computer Methods and Programs in Biomedicine 170 (2019) 95 – 106.

[17] S. Niroomandi, D. Gonzlez, I. Alfaro, F. Bordeu, A. Leygue, E. Cueto, F. Chinesta, Real-time simulation of biological soft tissues: a pgd approach, International Journal for Numerical Methods in Biomedical Engineering 29 (2013) 586–600.

[18] D. Borzacchiello, J. V. Aguado, F. Chinesta, Non-intrusive Sparse Subspace Learning for Parametrized Problems, Archives of Computational Methods in Engineering 26 (2019) 303–326.

[19] ©ANSYS, "Dynamic ROM Components." Twin Builder 2020R1 Online Help, Inc. 2020.

[20] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747 (2016).

[21] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (1997) 1735–1780.

[22] J. F. Kolen, S. C. Kremer, Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies, IEEE, pp. 237–243.

[23] N. Hermant, P. Perrier, Y. Payan, Human tongue biomechanical modeling, in: Y. Payan, J. Ohayon (Eds.), Biomechanics of Living Organs: Hyperelastic Constitutive Laws for Finite Element Modeling, London, UK: Academic Press, 2017, pp. 395–411.

[24] P. Hoole, A. Zierdt, Five-dimensional articulography, in: B. Maassen, P. van Lieshout (Eds.), Speech Motor Control: New developments in basic and applied research, Oxford, UK: Oxford University Press, 2010, pp. 331–349.