HHS Public Access

Author manuscript

Interspeech. Author manuscript; available in PMC 2023 December 15.

Published in final edited form as:

Interspeech. 2023 August ; 2023: 4189–4193. doi:10.21437/interspeech.2023-2247.

Motor control similarity between speakers saying “a souk” using inverse atlas tongue modeling

Ursa Maity[1], **Fangxu Xing**[2], **Jerry Prince**[3], **Maureen Stone**[4], **El Fakhri Georges**[2], **Jonghye Woo**[2], **Sidney Fels**[1]

[1]School of Biomedical Engineering, University of British Columbia, Canada

[2]Gordon Center for Medical Imaging, Harvard Medical School, Boston, USA

[3]Department of Electrical and Computer Engineering, John Hopkins University, USA

[4]University of Maryland Dental School, Baltimore, USA

Abstract

Finite element models (FEM) of the tongue have facilitated speech studies through analysis of internal muscle forces indirectly derived from imaging data. In this work, we build a uniform hexahedral FEM of a tongue atlas constructed from magnetic resonance imaging data of a healthy population. The FEM is driven by inverse internal tongue tissue kinematics of speakers temporally aligned and deformed into the same atlas space, while performing the speech task “a souk” allowing muscle activation predictions. This work aims to investigate the commonalities in tongue motor strategies in the articulation of “a souk” predicted by the inverse tongue atlas model. Our findings report variability among five speakers for estimated muscle activations with a similarity index using a dynamic time warp function. Two speakers show similarity index > 0.9 and two others < 0.7 with respect to a reference speaker for most tongue muscles. The relative motion tracking error of the model is less than 2% which is promising for speech study applications.

Index Terms

speech production; biomechanical modeling; tongue atlas

1. Introduction

The tongue is a highly deformable organ that plays a major role in speech production via interdigitated muscles activated with neuro-anatomical signals locally rather than as a whole unit [1]. In previous studies, the internal tongue deformation during specific speech utterances has been effectively captured using state-of-the-art medical imaging modalities to study speech synthesis [2]. However, there remains a gap in understanding how the biomechanical properties of the tongue musculature tie to speech synthesis using even simple motions such as the tongue’s protrusion or retraction [3]. Determining the relationship between the neuro-muscular activations and articulatory gestures in these

ursa.maity@ubc.ca .

speech utterances has been a long-standing issue in speech research areas [3–6], particularly those related to inter-subject variability of tongue anatomy and functionality of tongue muscles.

The aforementioned problem in speech research may be solved by investigating the range of motor functions of the tongue muscles in producing the same sounds with different speakers. Direct measurement strategies such as electromyography (EMG) for studying tongue muscle function are invasive and noisy [7]. On the other hand, medical imaging techniques such as magnetic resonance imaging (MRI), ultrasound, and computed tomography (CT) can provide highly detailed morphological information of oropharyngeal structures for studying deformations during sound production, chewing, or swallowing. However, they do not directly measure internal muscular forces. Finite element (FE) modeling and simulation of oropharyngeal structures has proved to be an effective strategy to quantify aspects of speech production, such as internal tissue forces and muscle activation patterns, which cannot be directly derived from medical imaging [8]. The modeling approaches for this purpose can be broadly classified into two types: 1) forward modeling, measuring tongue deformation from muscle activations and 2) inverse modeling, estimating muscle activation patterns from internal tongue deformation [9].

Recently, as related developments, Woo et al. [10] developed a 4D atlas of tongue motion using both cine and tagged MRI while uttering the words “a souk” and “a geese.” To establish a pattern of variability in a given population uttering these words, Xing et al. [11] developed a statistical approach using correlation among different internal muscles. Muscle activation correlation patterns for subjects were estimated using the tissue deformation fields obtained from tagged MRI of individual subjects morphed into the atlas space. However, this approach did not account for the nonlinear hyperelastic properties of the tongue tissue, which are often modeled in state-of-the-art FE models [4, 8]. Based on motion data in subject spaces from tagged MRI, Harandi et al. [7] developed subject-specific FE models for four subjects using the Artisynth toolkit [12] that predicted activations using a quadratic inverse solver [13]. However, the predictions were suboptimal [4] due to significant relative tracking error and the lack of internal muscle fiber directions of each speaker, which are difficult to acquire without invasive techniques.

To overcome these shortcomings, in this work, we propose a data-driven approach of inverse FE modeling to measure variability in muscle functions among different subjects by simulating subject-specific motion data [11] morphed into the statistical atlas geometry of healthy individuals. The FE model is built by registering the muscle fiber directions of a cadaver tongue [14] in the same atlas space. Specifically, we provide a quantitative analysis to show similarities in the motor strategies employed while speaking the same word “a souk” among five American English speakers using inverse tongue atlas modeling. Our proposed inverse atlas model can predict muscle functions for several subjects, thus saving the complex and time-consuming process of generating a subject-specific model for each speaker. Moreover, the model provides morphological structures that can be used to register bone geometry, and the predictions only depend on the motion deformation fields of each subject in the atlas space. We report the findings based on the estimated muscle activation patterns, including: 1) variation in muscle excitation among the five speakers for the same

vowel and consonant sounds, and 2) function of the tongue protruder and retractor muscles for aforementioned sounds.

2. Materials and Methods

2.1. Data Acquisition and Atlas Construction

The cine and tagged MRI data were acquired using a 12-channel head and a four-channel neck coil on a Siemens 3.0T Tim-Trio MRI scanner. The in-plane image resolution was 1.875mm × 1.875mm, and the slice thickness was 6 mm. The following sequence parameters were used: a repetition time (TR) of 36 ms, an echo time (TE) of 1.47 ms, a flip angle of 6, and a turbo factor of 11.

To obtain the displacement vectors of tissue points inside the tongue volume of a particular subject over time in their corresponding subject space, we utilized a phase vector incompressible registration algorithm (PVIRA) [11]. Specifically, PVIRA reconstructed a dense, 3D, and incompressible motion field at each time frame by tracking the corresponding harmonic phase data from tagged MR volumes. To construct the atlas, we applied a registration strategy that deformed the displacement vectors of the subjects speaking “a souk” and “a geese” into the average atlas space. The average atlas space was then determined by spatially aligning all healthy subjects. The phrases “a souk” and “a geese” as target utterances are of phonetic importance in speech studies, as it primarily involves superior-posterior and anterior-posterior tongue motion, minimal lateral motion, and limited jaw and lip movement. This particular study uses the motion data of two male and three female native English speakers in the age range of 20–45 years uttering “a souk”.

2.2. Tongue Atlas Model Design

2.2.1. Voxelized FEM generation—The multi-subject atlas is modeled as a voxelized volumetric mesh, comprising 6,166 nodes and 5,070 uniform hexahedral elements generated using the work by Lloyd et al. [12] to create embedded FEM in Artisynth. The algorithm generates a cuboid or a bounding FEM around any polygonal surface mesh. The number of elements of this bounding FEM that are more than 50% inside the surface mesh are used to generate the voxelized FEM as shown in Figure 1.a). For the purpose of visualization, a surface mesh generated from the atlas segmentation mask is added to the model.

2.2.2. Tongue Muscles definition—The following muscles are modeled in the tongue FEM, as shown in Figure 1.b): genioglossus (GG), hyoglossus (HG), styloglossus (STY), geniohyoid (GH), and myohyoid (MH), which are the extrinsic muscles (connected to the bone), and transverse (TRANS), verticalis (VERT), superior longitudinal (SL), inferior longitudinal (IL), which are intrinsic muscles (located inside the tongue). The genioglossus (GG), transverse (TRANS), and vertical (VERT) muscle bundles are further divided into five smaller functionally relevant segments (a: posterior to e: anterior), to accommodate more degrees of freedom [5, 15]. Using the same algorithm in Section 2.2.1 to generate a bounding FEM around each segmented muscle mask from structural MRI [11], the hexahedral elements in the total FEM volume are allocated to different muscle bundles. Next, we define the resting fiber directions inside each element. However, obtaining accurate

information on fiber directions from structural MRI is challenging. A viable approach is to register the muscle fibers defined in state-of-the-art FE models [4], which have fiber orientations based on a cadaver tongue [8].

In order to obtain the muscle fiber directions for the corresponding elements for each muscle bundle, a two-step approach of mesh registration and deformation is carried out. The registration is carried out via Iterative Closest Point Mesh Correspondence with Gaussian weight function parameters. The maximum weight was set to 1, and the standard deviation was set to 0.01. The dynamic registration controller used is described in detail by Khallaghi et al. [16]. The surface mesh (assigned as the source) is taken from a model that contains the aforementioned muscle fiber directions and a voxelized FEM is generated for that mesh in the same way as mentioned in Section 2.2.1. This voxelized FEM deforms around the atlas surface mesh, which serves as the target. The deformed voxelized FEM is then used to perform a non-linear geometry transformation of the muscle fibers for each muscle segment in the selected subject-specific model. The geometry transformation in Artisynth developed by Lloyd et al. [17] is carried out using a piece-wise smooth deformation field from the deformed FEM. The transformer that generates an affine transformation which is applied to the muscle fibers is described in Section 3.4 of the Artisynth toolkit [12]. The deformed and transformed muscle fiber directions are used to create a Delaunay interpolation matrix that calculates the muscle fiber direction at rest for each hex element of the Atlas FEM for every corresponding muscle bundle [17].

2.2.3. Material Properties—The voxelized Tongue Atlas FEM is modeled using a non-linear hyperelastic material similar to the models by Harandi et al. [4]. The FE model uses a fifth-order Mooney-Rivlin Material, and we refer the reader to Buchaillard et al.'s work [8] for the definition of the material parameters. We replicate the strain energy modifications done by Harandi et al. [7] in their FE model and define the muscle bundle material as a Blemker muscle [18]. Further details on how the parameters are chosen can be found in the work by Harandi et al. [19].

2.3. Data-driven Inverse simulation

Using Stavness et al.'s inverse tracking controller [13], we predict muscle activation patterns of five speakers saying "a souk." The controller utilizes the deformed internal tissue kinematics of each speaker in the atlas space, with linearly interpolated displacement vectors from the motion field on selected FEM nodes serving as input. The simulation of articulatory trajectory relies on muscle redundancy (many-to-one mapping) and imposed constraints. The estimated muscle activations can either reflect realistic patterns or highlight limitations in the model based on existing anatomical knowledge.

2.3.1. Target Nodes selection—Selection of the target nodes affects the simulations and we address it as follows: a) we select regions with high velocity densities, b) nodes are selected along the muscles that are expected to get activated the most. For this study, we select four clusters of five nodes each, extending from the tip to the posterior end of the central muscles in the tongue FEM.

2.3.2. Inverse tracking controller—The inverse tracking controller solves for normalized activation values a for muscle exciter terms by minimizing the loss term as given by

$$\arg \min_{a} \left(w_m \| v - H_m a \|^2 + \frac{\alpha}{2} a^T a \right), \quad (1)$$

where $w_m \| v - H_m a \|^2$ is the velocity tracking error term of the 20 target points, and H_m is a matrix summarizing the biomechanical characteristics of the system, such as mass, joint constraints, and force-activation properties of the muscles. $\frac{\alpha}{2} a^T a$ is the l^2-norm regularization term to reduce muscle redundancy and distribute the predictions among the exciters equally. The weight of the motion target component w_M was set to 1.1, and the regularization coefficient α was set to 0.01 to achieve the lowest possible tracking error, while maintaining model stability.

2.3.3. Error Analysis—We evaluate the model from a data-driven perspective, by assessing its ability to accurately reproduce the actual tongue tissue displacement motion observed in tagged MRI. To measure the performance of the inverse model, we calculate the Root Relative Mean Square Error (RRMSE) and the Relative Absolute error (RAE) for each time frame of the simulation as given by

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{3n} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{3n} \hat{x}_i^2}}, \; RAE = \sqrt{\frac{\sum_{i=1}^{3n} |x_i - \hat{x}_i|}{\sum_{i=1}^{3n} |x_i - \bar{x}|}}, \quad (2)$$

where x_i represents the actual displacement of each target node, while $\hat{x}_i$ is the target displacement in any of the three directions of tongue movement, namely forward, upwards, or lateral. The total number of target nodes is denoted by n, and the sum of errors is calculated for $3n$ directions (forward, upward, and lateral). The mean of actual displacements of all nodes in each time frame is represented by $\bar{x}$. We report both the maximum and mean error values across the time frames of motion simulation.

2.4. Similarity Measurement

The input tagged MRI data in the atlas space for each subject were initially recorded at a rate of 26 frames per second, with each frame lasting 38.46ms. Of note, for a smooth simulation, this data is interpolated across 50 time frames in the inverse model. Although all speakers perform the speech task “a souk” in the same duration, there are slight variations in the time frames of the utterances for /ə/, /s/, /u/, and /k/ for each speaker, and these are identified visually by a speech scientist from the sagittal view of cine MRI [7]. To incorporate this variation into our quantitative analysis of the similarities in muscle motor strategies for the five speakers using the inverse tongue atlas model, we apply a Dynamic Time Warping (DTW) algorithm [20], which is commonly employed for speech recognition. The DTW function calculates the minimum sum of Euclidean distance d, between estimated muscle activation curves for two subjects for any given muscle exciter (see Figure 2).

As the value of d increases, the difference between the two curves also increases. The Euclidean sum d_{ij} is calculated for each exciter i across all subjects j, resulting in the matrix $D_{p\times m}$ as given by

$$D_{p\times m}=\begin{bmatrix} d_{11} & \dots & d_{1m} \\ d_{21} & \dots & d_{2m} \\ \cdot & \dots & \cdot \\ d_{p1} & .. & d_{pm} \end{bmatrix}, \tag{3}$$

where p is the number of exciters, and m is the number of subjects. The similarity index matrix $S_{p\times m}$ is defined by its elements s_{ij}, which are calculated through normalization, as follows:

$$s_{ij}=1-\frac{|d_{ij}-d_{min}|}{|d_{max}-d_{min}|}, \tag{4}$$

where d_{max} and d_{min} are the maximum and minimum values of d_{ij}, respectively. The value of s_{ij} ranges from 0 to 1, with 0 indicating no similarity between signals and 1 indicating identical signals. For our evaluation, we use speaker S1 as the control, meaning that each element in the first column of the matrix S is set to 1.

3. Results

3.1. Tongue Muscles activation for “a souk”

This section interprets the muscle activation patterns. The middle and anterior fibers of the genioglossus (GGM/GGA) lower the dorsum of the tongue, causing an overall backward motion [21]. Geniohyoid (GH) and myohyoid (MH) are floor muscles that assist in tongue elevation and protrusion [21]. Additionally, it is known that extrinsic muscles styloglossus (STY) and hyoglossus (HG) aid in backward/upward and backward/downward tongue motion, respectively. Intrinsic muscles superior-longitudinal (SL) and inferior-longitudinal (IL) both aid in tongue retraction and also elevate/lower the tongue-tip, respectively [21]. The estimated muscle activations are plausible indicators, not actual values, revealing potential muscle activities aligned with predictions and identifying similarities in motor strategies for “a souk.”

3.1.1. Articulation for /ə/ to /s/—GGM is significantly activated (except for S4) during /ə/, while GGA shows slight activation. As the tongue elevates and pushes forward for /s/, distinct muscle activations occur. VERT segments are mostly inactive during /ə/, except for minor activation in VERTe. However, VERTa-b gradually increases activation during /s/. TRANSe is active throughout /ə/, while TRANS(a-c) gradually activate during /s/. HG and SL muscles exhibit minimal activation initially to aid in slight retraction and tip elevation for /ə/.

3.1.2. Transition from /s/ to /u/—While uttering /s/ and transitioning to the high vowel /u/, the tongue holds a constrained pose for a few milliseconds and proceeds to move forward. As such, it is expected that the tongue protrudor muscles will get increasingly activated while there will be a decline in activation of tongue retractor muscles. Thus, it

makes complete sense to observe a slow rise in activation of TRANSa-b and MH and slow decline in activation of STY and GH in the predicted pattern. TRANSa and VERTb exhibit peak activation, while GGPa shows the lowest activation near the production of the high vowel /u/. The tongue retractor muscles, such as GGM, GGA, VERTe, and HG exhibit near-zero activation levels.

3.1.3. Transition from /u/ to /k/—Lastly, while uttering /k/, the tongue tip slowly moves downward, but remains relatively close to the position it maintains during the /u/ sound. Hence, the only difference observed from the previous transition period is a further steady decline in muscles that elevate the tongue tip and push the tongue forward. The observations are consistent with these predictions, as observed with a slight decrease in activation of TRANSa-b. In all the speakers, there is also a slight increase in activation towards the end of /k/ with some assistance from tongue retractor muscles, including TRANSe and GGPa.

3.2. Tracking Error

Table 1 shows both the maximum and average RRMSE and RAE values (expressed in percentage) for five speakers over the time period of 1s simulated at 50 fps.

3.3. Similarity Index using Dynamic Time Warp

The values of the similarity index matrix $S_{m\times p}$ (described in Section 2.4) using the DTW function are calculated for each speaker (with respect to S1) for 21 muscle exciters and represented as a heatmap as in Figure 3. The similarity index between S1 and S2 is very high (over 0.9) for seven exciters, moderately high (between 0.7–0.9) for eight exciters, and very low (below 0.2) for two exciters. Between S1 and S3, ten exciters have a similarity index >0.9. For S4, it is <0.5 for seven exciters and >0.9 only for five exciters indicating low similarity with S1. Lastly, for S5, it is > 0.9 for eight exciters, between 0.7–0.9 for five exciters, and <0.5 for five exciters.

4. Discussion and Conclusion

This study used the inverse atlas tongue to estimate muscle activation patterns for five native English speakers saying the word “a souk.” The results showed that tongue retractor muscles were observed to be most activated during the utterance of /ə/, and the tongue protrudor muscles were relatively more active during the utterance of /s/ and peaked around the high vowel /u/ with steady decline around /k/ for most speakers. We defined a Similarity Index using a DTW function to quantify the similarities in the estimated muscle activation patterns of the five speakers. These values indicate there are common motor strategies that may be explored by analyzing these plausible muscle activations. However, there are a few limitations in this work, including 1) the muscle fiber orientations are assumed to be the same in the multi-subject atlas for all five speakers, 2) the FE model has limited degrees of freedom, and 3) the inverse tracking controller can track only a definite number of nodes accurately so far. In future work, we will analyze motor strategies in the production of ”a souk” and ”a geese” based on observed muscle activation patterns (Section 3.3) and acoustic

formant analysis. We plan to expand the study to include more speaker datasets and examine muscle activation during ”a geese.”

Acknowledgements

This work was supported by NIH grant R01DC018511 and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1]. Stone M, Woo J, Lee J, Poole T, Seagraves A, Chung M, Kim E, Murano EZ, Prince JL, and Blemker SS, “Structure and variability in human tongue muscle anatomy,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 6, no. 5, pp. 499–507, 2018. [PubMed: 30135746]

[2]. Badin P, Bailly G, Reveret L, Baciu M, Segebarth C, and Savariaux C, “Three-dimensional linear articulatory modeling of tongue, lips and face, based on mri and video images,” Journal of Phonetics, vol. 30, no. 3, pp. 533–553, 2002.

[3]. Kappert K, van Alphen M, van Dijk S, Smeele L, Balm A, and van der Heijden F, “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, vol. 22, no. 8, pp. 827–839, 2019. [PubMed: 30963800]

[4]. Harandi NM, Woo J, Farazi MR, Stavness L, Stone M, Fels S, and Abugharbieh R, “Subject-specific biomechanical modelling of the oropharynx with application to speech production,” in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE, 2015, pp. 1389–1392.

[5]. Stone M, Epstein MA, and Iskarous K, “Functional segments in tongue movement,” Clinical linguistics & phonetics, vol. 18, no. 6–8, pp. 507–521, 2004. [PubMed: 15573487]

[6]. Perrier P, Payan Y, Zandipour M, and Perkell J, “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. [PubMed: 14514212]

[7]. Harandi NM, Woo J, Stone M, Abugharbieh R, and Fels S, “Variability in muscle activation of simple speech motions: A biomechanical modeling approach,” The Journal of the Acoustical Society of America, vol. 141, no. 4, pp. 2579–2590, 2017. [PubMed: 28464688]

[8]. Buchaillard S, Perrier P, and Payan Y, “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. [PubMed: 19813813]

[9]. Stavness I, Lloyd JE, Payan Y, and Fels S, “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.

[10]. Woo J, Xing F, Stone M, Green J, Reese TG, Brady TJ, Wedeen VJ, Prince JL, and El Fakhri G, “Speech map: A statistical multimodal atlas of 4d tongue motion during speech from tagged and cine mr images,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 7, no. 4, pp. 361–373, 2019. [PubMed: 31328049]

[11]. Xing F, Stone M, Goldsmith T, Prince JL, El Fakhri G, and Woo J, “Atlas-based tongue muscle correlation analysis from tagged and high-resolution magnetic resonance imaging,” Journal of Speech, Language, and Hearing Research, vol. 62, no. 7, pp. 2258–2269, 2019.

[12]. Lloyd JE, Stavness I, and Fels S, “Artisynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation,” in Soft tissue biomechanical modeling for computer assisted surgery. Springer, 2012, pp. 355–394.

[13]. Stavness I, Lloyd JE, and Fels S, “Automatic prediction of tongue muscle activations using a finite element model,” Journal of biomechanics, vol. 45, no. 16, pp. 2841–2848, 2012. [PubMed: 23021611]

[14]. Gérard J-M, Wilhelms-Tricarico R, Perrier P, and Payan Y, “A 3d dynamical biomechanical tongue model to study speech motor control,” arXiv preprint physics/0606148, 2006.

[15]. Miyawaki K, “A preliminary report on the electromyographic study of the activity of lingual muscles,” Ann Bull RILP, vol. 9, pp. 91–106, 1975.

[16]. Khallaghi S, Sánchez CA, Rasoulian A, Sun Y, Imani F, Khojaste A, Goksel O, Romagnoli C, Abdi H, Chang S et al. , “Biomechanically constrained surface registration: Application to mr-trus fusion for prostate interventions,” IEEE transactions on medical imaging, vol. 34, no. 11, pp. 2404–2414, 2015. [PubMed: 26054062]

[17]. Lloyd JE, Sánchez A, Widing E, Stavness I, Fels S, Niroomandi S, Perrier A, Payan Y, and Perrier P, “New techniques for combined fem-multibody anatomical simulation,” in New Developments on Computational Methods and Imaging in Biomechanics and Biomedical Engineering. Springer, 2019, pp. 75–92.

[18]. Blemker SS, Pinsky PM, and Delp SL, “A 3d model of muscle reveals the causes of nonuniform strains in the biceps brachii,” Journal of biomechanics, vol. 38, no. 4, pp. 657–665, 2005. [PubMed: 15713285]

[19]. Harandi N, “3d subject-specific biomechanical modeling and simulation of the oral region and airway with application to speech production,” Ph.D. dissertation, University of British Columbia, 2016.

[20]. Sakoe H and Chiba S, “Dynamic programming algorithm optimization for spoken word recognition,” IEEE transactions on acoustics, speech, and signal processing, vol. 26, no. 1, pp. 43–49, 1978.

[21]. Zemlin WR, “Speech and hearing science, anatomy and physiology.” 1968.

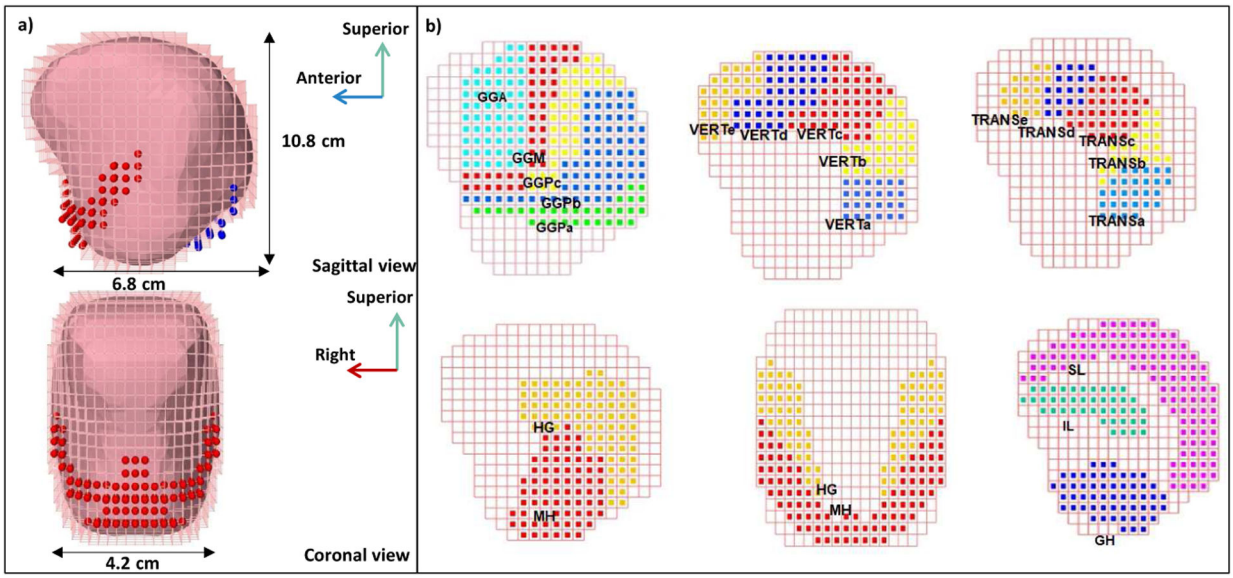

Figure 1:

a) Sagittal and coronal perspective view of the atlas tongue FEM. The attachment of the tongue to the jaw and hyoid bones are also shown in red and blue respectively b) Muscle bundles defined in the hexahedral mesh. Top row: central muscles GG, VERT and TRANS divided into 5 distinct segments (a: posterior to e: anterior) in mid-sagittal view. Bottom row shows the muscle segments HG, MH, SL, IL and GH. A midcoronal view (bottom center) is also included for HG and MH.

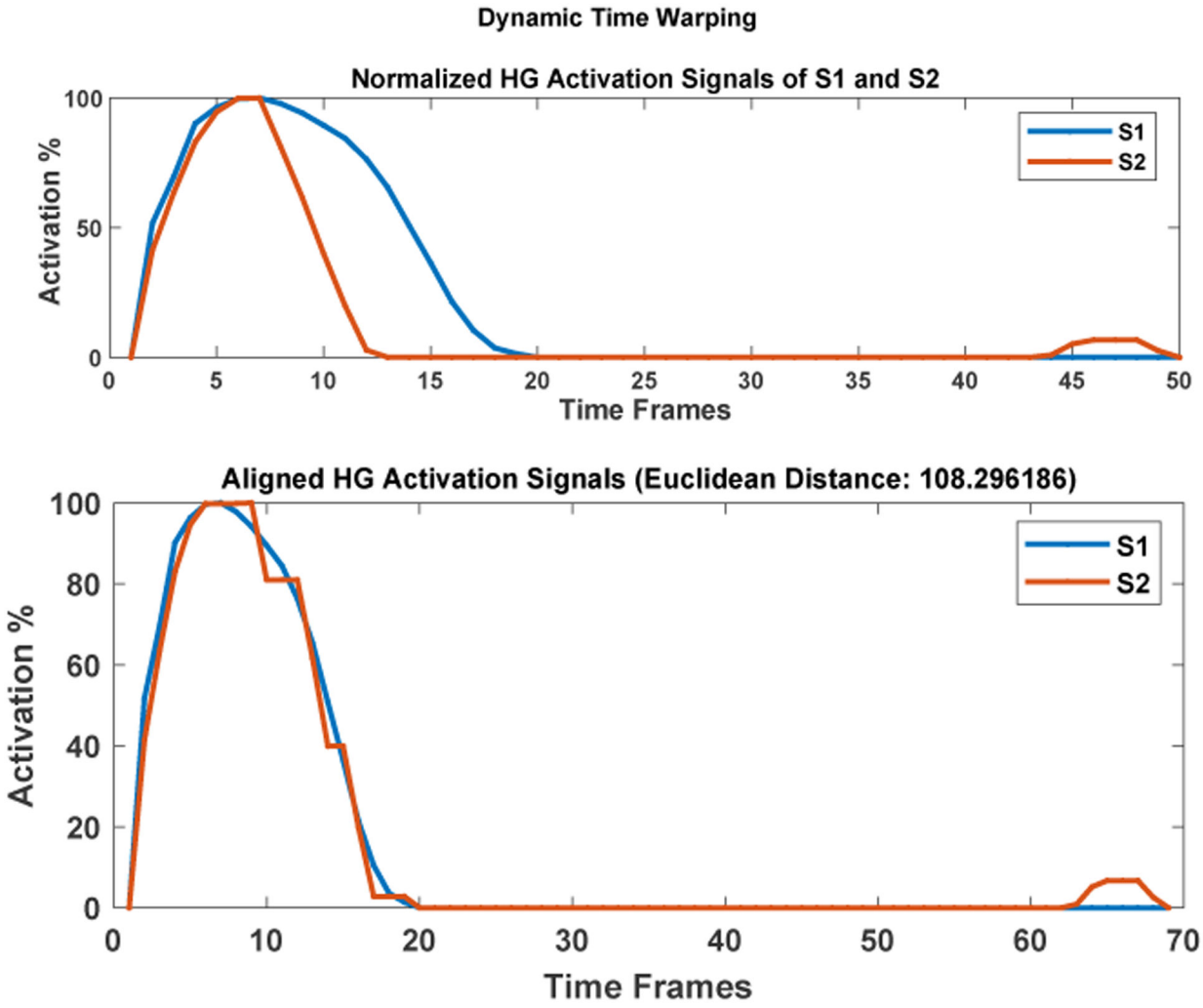

Figure 2:
Dynamic Time Warping to calculate d_{ij} for muscle HG between S1 and S2

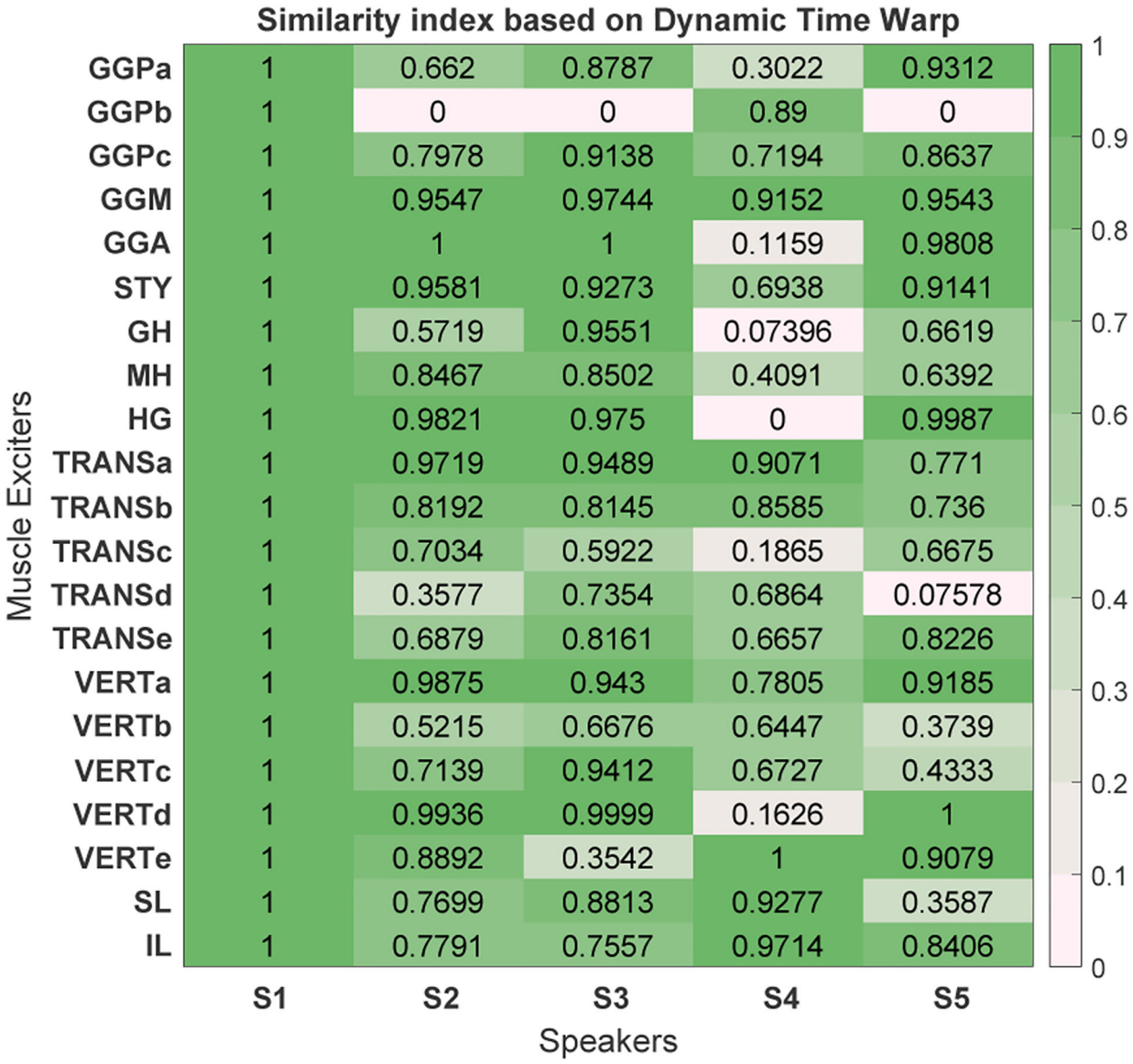

Figure 3:
Similarity Index between S1 and the rest of speakers based on DTW. Darker green indicates a higher level of similarity.

Table 1:

Error Analysis

		S1	S2	S3	S4	S5
RAE%	Mean	0.381	0.873	0.672	0.575	0.535
	Max	0.567	1.521	1.020	1.079	0.805
RRMSE%	Mean	0.044	0.103	0.077	0.069	0.063
	Max	0.068	0.180	0.111	0.131	0.098