3D AUDITORY-ARTICULATORY MODELING OF THE LARYNGEAL CONSTRICTOR MECHANISM

Scott R. Moisik and John H. Esling

Department of Linguistics, University of Victoria, Victoria, BC, Canada
srmoisik@uvic.ca, esling@uvic.ca

ABSTRACT

The vocal tract is reinterpreted in the context of the laryngeal articulator model, which integrates the functions of the laryngeal and oral components of the vocal tract. To account for the action of pharyngeal-resonator reduction, for constricted phonation types, and for the interaction of glottal pitch with the laryngeal constrictor mechanism, a three-dimensional model has been developed on the basis of auditory parameters and extrapolation from articulatory data sources. A critical aspect of the proposed model is the functioning of the aryepiglottic sphincter, formed by the aryepiglottic folds at the upper border of the larynx articulating towards the epiglottis, in such sounds as [ʡ, ʢ, ʜ, ʕ, ħ]. The novel feature of this model is the inclusion of a separate and 'reversed' action of the laryngeal component.

Keywords: laryngeal articulator, constrictor, vocal tract, pharyngeal resonator, aryepiglottic sphincter.

1. INTRODUCTION

An auditory-articulatory model of the vocal tract must achieve specific goals if it is to accurately represent the production of sounds generated within the pharynx, under the control of the laryngeal constrictor mechanism [7]:

- it must account for complete stricture at the aryepiglotto-epiglottal place of articulation, i.e. full airway closure in an epiglottal stop [ʡ];
- it must contain gradations away from full stricture to account for approximant [ʕ] and fricative [ħ] articulations;
- it must allow for the presence of aryepiglottic fold trilling to account for the production of the pharyngeal/epiglottal trills [ʢ] and [ʜ];
- it must include an effect on the lower, medial parts of the laryngeal mechanism to account for the less extreme stricture required to stop the glottal vocal folds in the production of [ʔ];
- it must emulate the massing of the ventricular folds with the vocal folds not only during stopped [ʔ] but also during phonation types that result from higher degrees of constriction, such as creaky voice, whispery phonation, and different degrees of harsh voice along a scale of low to high pitch.

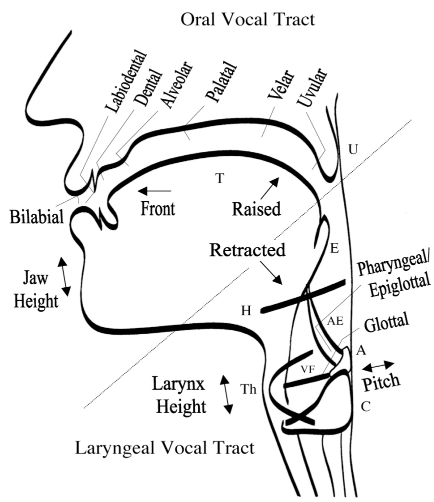
The fundamental categories on which these criteria are based have been described in a revised set of phonetic states of the glottis [9], which have also been referred to as states of the larynx. The theoretical model itself has been proposed to account for a range of syllabic tonal register and pharyngeal consonantal distinctions that occur in various language families [5, 8].

Previous vocal tract models provide only a reductive treatment of the pharyngeal/laryngeal component [1, 2, 6]. Kahrilas et al.'s [11] 3D swallowing model illustrates the fundamental structural dynamics of laryngeal sphinctering, however, none of the soft tissues essential to the range and quality of the linguistic use of this structure are represented. The present model focuses on the linguistic, anatomical, and physiological gaps left by previous 3D models.

2. A TWO-VOCAL-TRACT 3D MODEL

There are pragmatic difficulties in visualizing the pharynx and larynx as articulation is taking place. The aryepiglottic folds, important to the formation of articulatory dynamics of pharyngeals and epiglottals, are poorly visualized in x-ray imaging techniques [14] while MRI requires lengthy imaging times and is subject to signal-to-noise ratio degradation when movement occurs [3]. These problems are compounded by the fact that the folds are delicate structures with an oblique trajectory through the hypopharyngeal space. Thus, the most common gestures argued to be active for these sounds are pharyngeal lumen compression, tongue root and epiglottis retraction, and laryngeal elevation. Laryngoscopy [5, 8, 12] is an attractive tool for the visualization of the aryepiglottic folds during the production of these sounds in addition to providing indices for laryngeal elevation and tongue root retraction. Based on such evidence, the vocal tract has been reconceptualized into two primary articulatory domains in the 'Laryngeal Articulator Model' [9, 5]. Fig. 1 illustrates this conceptual expansion of the articulatory capacity of the laryngeal vocal tract and the primacy of the aryepiglottic folds in the production of sounds that have been termed pharyngeals and epiglottals.

Figure 1: The 'two-vocal-tract' reconceptualization of the standard articulatory model of the vocal tract [9].



The 3D model developed in the research presented here represents the laryngeal section of the vocal tract, which is conceived of as a series of valves, each of which is responsible for a range of articulatory configurations and is made up of its own unique anatomical structures.

Table 1: Pharyngeal and laryngeal valves [5] and their anatomical identity.

Valve Name	Anatomical Structure
V1	vocal folds
V2	ventricular folds
V3	aryepiglottic folds
V4	tongue root / epiglottis
V5	laryngeal height
V6	pharyngeal walls

The 3D articulatory model focuses on these six basic valves and their movement capabilities and contingencies. The odd-numbered valves are dominant in their relation to the even-numbered valves. Dominant valves are unique in that they are physiologically independent of the subordinate valves. For example, it is impossible to engage the ventricular folds (V2) without first adducting the vocal folds (V1) and it is impossible to retract the tongue root and epiglottis (V4) without first engaging the aryepiglottic folds (V3) [5, 15].

2.1. Implementation of the theoretical model into the design of the 3D model

Each valve has a set of parameters that specify its unique physiological capabilities. The type of linguistic output possible with any given valve is contingent upon its parameter settings. The 3D model design is based on these movement parameters, which relate directly to the individual muscles responsible for the movement possibilities of each valve. Users will be able to manipulate these parameters through a Graphical User Interface (GUI) interface with adjustable sliders (illustrated in Figs. 2 and 3) to observe each setting's physiological and linguistic relevance.

2.2. Model Design

The model is a general-purpose tool for the study and analysis of pharyngeal/laryngeal articulations and their physiological properties. Each layer in the anatomical representation details a specific set of related tissues and can be viewed independently of the other layers. This feature allows the user to flexibly visualize the physiological behaviour of anatomical structures during articulation. The three main layers are: cartilage and ligament, musculature, and soft tissue. In addition to parameter control, the GUI allows the user to determine the tissue layers that appear on-screen and the global visualization of the model. The model can be freely viewed or sectioned from any angle or plane.

3. MODELING METHODOLOGY

The model polygon meshes were constructed using *Blender* (version 2.42a: win32 edition). The model is based on a composite body of data drawn from radiological studies of the larynx and hypopharynx as well as linguistic imaging studies and other anatomical resources. A significant resource is Hirano & Sato's histological atlas of the larynx [10]. Wherever possible, measurement averages were implemented to optimize the anatomical accuracy of the model; many of these averages have been derived from Eckel et al. [4] (see Table 2 for an example of the measurement averages used in this model).

The following process was used in constructing the model:

- Anatomical data relevant to the structure being represented were visually aligned and scaled in the modeling software;
- For symmetrical structures, one-half of the structure was traced using the modeling software; the resulting trace was mirrored and merged with its counterpart;
- Mesh construction was oriented about the axial plane; traces were stacked up according to position in vertical space;
- Vertices from the traces were manually linked to form mesh surfaces;
- Subsurfaces determined by a Catmull-Clark subdivision algorithm [13] were applied to the polygon mesh to give the model a smooth look.

Table 2: Anatomical measurement averages (from male larynges) implemented in the model: vocal folds and some related structures.

Anatomical Structure	Measurement (mm)
Vocal folds: length	13.8mm
Vocal folds: width	4.2 mm
interarytenoidal gap	6.2 mm
arytenoid cartilage *(musc.proc.-apex)*	17.1 mm
arytenoid cartilage *(voc.proc.-apex)*	17.0 mm
arytenoid cartilage *(musc.proc.-voc.proc.)*	13.5 mm

3.1. Reconstruction of laryngeal movements

Movement parameters for the model were largely determined by an analysis of variation in movement patterns during the production of various laryngeal and pharyngeal articulations. Canonical production of the following articulations were recorded laryngoscopically: breathing, aspiration, modal voice, glottal stop, epiglottal stop, whisper, breathy voice, whispery voice, creaky voice, falsetto, harsh voice (high, mid, low), voiced and voiceless pharyngeal fricatives, and voiced and voiceless epiglottal trills. Estimations of movement range were determined by comparing articulatory adjustments for a set of predefined parameters with a default setting of the larynx. The default setting was determined to be the posture under normal conditions of quiet breathing. Table 3 lists the movement parameters identified and their movement ranges. Fig. 4 illustrates how these measurements were obtained.

Table 3: Primary movement parameters determined by laryngoscopic imaging. Each anatomical parameter is conceived of as an aperture, which can be variably set for degree of openness (where 0% = minimal aperture; 100% = maximal aperture).

Articulation	CA	CC	IA-ET	VF
modal voice	8%	30%	48%	60%
creaky voice	0%	25%	22%	40%
breathy voice	0%	39%	63%	83%
whispery voice	5%	30%	49%	33%
falsetto	0%	22%	70%	60%
whisper	8%	27%	43%	17%
harsh voice (low)	3%	9%	0%	12%
harsh voice (mid)	6%	10%	20%	12%
harsh voice (high)	0%	14%	43%	15%
aspiration	25%	47%	88%	68%
glottal stop	0%	24%	25%	0%
epiglottal stop	0%	0%	0%	0%

CA = corniculate–arytenoid tubercles; CC = cuneiform cartilages; IA–ET = interarytenoid notch – epiglottic tubercle; VF = ventricular folds

Aperture percentages are used to define mesh deformation parameters for each of the specific anatomical landmarks referred to in Table 2. For CA, CC, and VF, movement is towards the sagittal midline of the larynx (defined by the glottis). The only exception to this is IA-ET, which requires structural approximation along an antero-posterior vector. In all cases, equal movement gradients were assumed.

3.2. Model presentation

The GUI provides the user with controls to manipulate the model along individual muscle constriction parameters (where commensurate contraction of muscle pairs is assumed). Predefined contraction patterns (referred to as ‘articulatory presets’) amongst groups of muscles under certain articulatory conditions (as outlined in Table 3) are made available to the user. Movement is synchronized with prerecorded audio of the articulation to illustrate the auditory effect of a gesture. Stills of the GUI illustrating the movement dynamics of high pitch harsh voice are presented in Figs. 2 and 3. These images show the laryngeal vestibule viewed from above, with the epiglottis situated at the bottom of the image and the cuneiform and corniculate tubercles at the top.

4. MODEL RELEVANCE

The model will serve as a tool to simulate articulation and physiological aspects of the pharynx and larynx, a region that is difficult to visualize. The model is extensible. Future additions may include: aerodynamic modeling, speech synthesis, and laryngeal embryology/ development.

Figure 2: Still from the 3D model interface showing the initial frame of the articulatory preset position preceding harsh voice (high pitch). Muscle contraction sliders allow for visualization of muscle activity during articulation. The circled slider indicates activity of the PCAs to maintain glottal aperture consistent with an unconstricted open laryngeal setting.

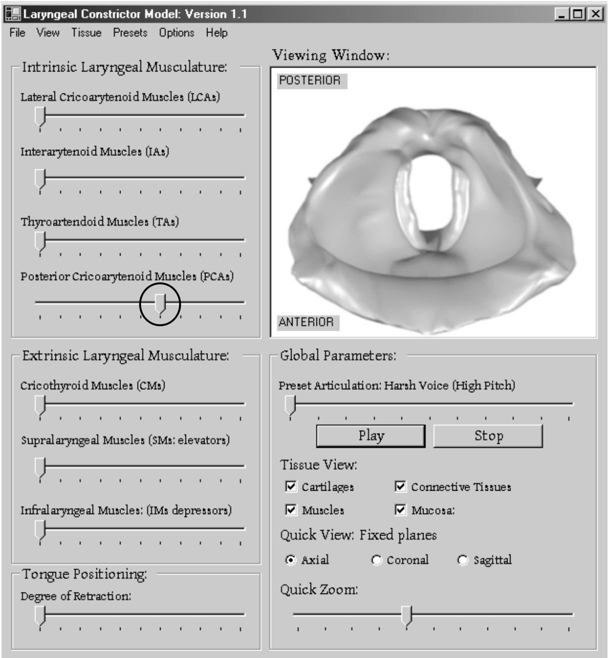


Figure 3: Still from the 3D model interface showing a final frame of the harsh voice (high pitch) articulation. The sliders which index the movement parameters of this constricted and stretched setting have been circled.

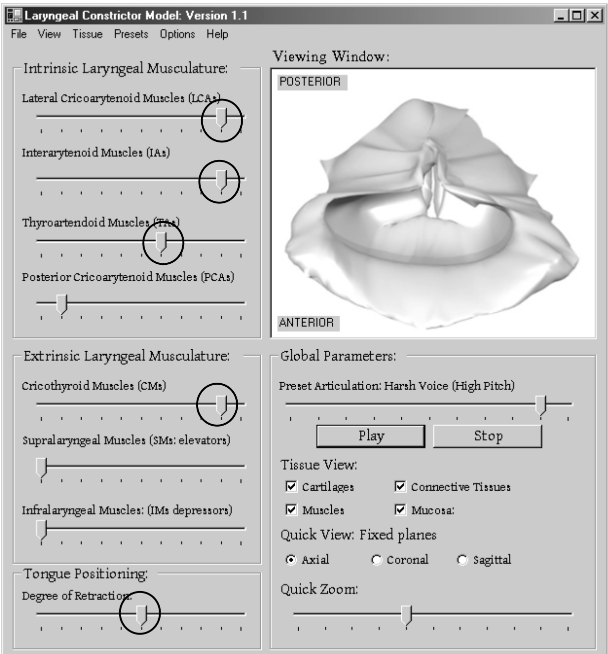
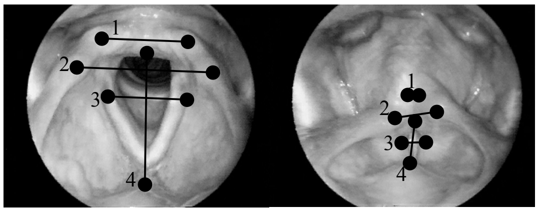


Figure 4: Two frames taken from a laryngoscopic sequence of harsh voice (high pitch) articulation. The left image shows an unconstricted posture (cf. Fig. 3), while the right image shows a constricted posture. Measurements for Table 3 were obtained from laryngoscopic stills such as these. Ruler-based measurement was done using Adobe Photoshop Elements. Reference points for the four apertures have been indicated with lines: 1 = corniculate – arytenoid tubercles; 2 = cuneiform cartilages; 3 = ventricular folds; 4 = interarytenoid notch – epiglottic tubercle.



5. REFERENCES

[1] Badin, P., Borel, P., Bailly, G., Reveret, M., Baciu, M., Segebarth, C. 2000. Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images. *5th Speech Production Seminar*, München.

[2] Birkholz, P., Jackel, D. 2003. A three-dimensional model of the vocal tract for speech synthesis. *Proc. 15th ICPhS*. 2597-2600. Barcelona.

[3] Curtin, H.D. 1989. Imaging of the larynx: current concepts. *Radiology* 173, 1, 3-11.

[4] Eckel, H.E., Sittel, C., Zorowka, P., Jerke, A. 1994. Dimensions of the laryngeal framework in adults. *Surgical Radiological Anatomy* 16, 31-36.

[5] Edmondson, J., Esling, J. 2006. The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology* 23, 157-191.

[6] Engwall, O. 1999. Vocal tract modeling in 3D. *KTH TMH-QPSR* 1-2/1999, 31-38.

[7] Esling, J. 1996. Pharyngeal consonants and the aryepiglottic sphincter. *JIPA* 26, 65-88.

[8] Esling, J. 1999. Voice quality settings of the pharynx. *Proc. 14th ICPhS* San Francisco, 2449-2452.

[9] Esling, J. 2005. There are no back vowels: the laryngeal articulator model. *Canadian Journal of Ling.* 50, 13-44.

[10] Hirano, M., Sato, K. 1993. Histological color atlas of the human larynx. San Diego: Singular Publishing Group.

[11] Kahrilas, P.J., Lin S., Chen, J., Logemann, J.A. 1995. Three-dimensional modeling of the oropharynx during swallowing. *Radiology* 194, 575-579.

[12] Painter, C. 1986. The laryngeal vestibule and voice quality. *Arch. of Oto-Rhino-Laryngology* 243, 329-337.

[13] Qin, H., Mandal, C., Vemuri, B.C. 1998. Dynamic Catmull-Clark subdivision surfaces. *IEEE Transactions on Visualization and Computer Graphics* 4(3), 215-229.

[14] Whalen, D.H. 2004. Measurements and modeling in speech production research. In: *Proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication*. MIT, 107-117.

[15] Zemlin, W.R. 1998. *Speech and Hearing Science: Anatomy and Physiology* (4 ed). Boston: Allyn & Bacon.