THE PEDAGOGY OF THE LARYNGEAL ARTICULATOR MODEL

John Esling[1], Allison Benner[1], Silvia Calamai[2], Chiara Celata[3], Lise Crevier Buchman[4], Míša Hejná[5], Scott Moisik[6]

[1]University of Victoria, [2]Università di Siena, [3]Università di Urbino, [4]Université Sorbonne Nouvelle & Université Paris-Saclay, Hôpital Foch, [5]Aarhus University, [6]Nanyang Technological University
[1]esling@uvic.ca, [1]abenner@uvic.ca, [2]silvia.calamai@unisi.it, [3]chiara.celata@uniurb.it, [4]lise.buchman1@gmail.com, [5]misa.hejna@cc.au.dk, [6]scott.moisik@ntu.edu.sg

ABSTRACT

The Laryngeal Articulator Model (LAM) approach to the vocal tract in speech production has emerged over the past decades as an alternative to a strictly glotto-lingual approach. The LAM expands explanatory power by acknowledging: 1. that the larynx is a complex articulator consisting of a network of structures; and 2. that articulations in the lower vocal tract interact with articulations in the oral vocal tract. The LAM approach to speech production is not yet widely implemented in phonetics and phonology teaching materials. The main goal of this paper is to explore the different challenges of laryngeal pedagogy in order to make it easier for scholars in the field to engage with the LAM in pedagogical contexts. To do this we present the authors' specific experiences using the LAM in different pedagogical situations: teaching voice quality, training research/teaching assistants, writing textbooks, training speech clinicians, and modelling speech production.

Keywords: Laryngeal Articulator Model, voice quality, pedagogy, clinical phonetics, modelling

1. INTRODUCTION

This paper explores some of the challenges of laryngeal pedagogy, focusing on the Laryngeal Articulator Model (LAM) [15]. We present the authors' experiences using the LAM in the contexts of teaching voice quality, training research/teaching assistants, writing textbooks, training speech clinicians, and modelling speech production.

2. TRAINING RESEARCH ASSISTANTS

The LAM increases the number of phonation types and states of the larynx for phonetic description and instruction, both in terms of their articulatory production and their auditory correlates. For example, various types of harsh voice, aryepiglottic (AE) trilling, whispery voice and raised larynx voice and the states glottal stop, prephonation and epiglottal stop have revised articulatory definitions. Each type/state in the model represents a stage or level along the continuum in the functioning of the aryepiglottic (AE) sphincter mechanism: from wide open to firmly shut. The most basic auditory distinction dividing this continuum is between "constricted" and "unconstricted" qualities. Research assistants (RAs) coding laryngeal utterances in language or infant-speech databases can be trained to distinguish relatively tight (whispery, creaky, harsh, AE-trilled, pressed) vs. relatively open (breathy, modal, falsetto) postures of the laryngeal mechanism. These postures/qualities represent a larger and more visualizable action as a more 3D concept than if the larynx were only an opening-closing glottis. Here we highlight the application of the LAM in training RAs, using the Infant Speech Acquisition (InSpA) project as an example [4, 13-14].

The InSpA project investigated the incidence and role of laryngeal constriction in infant speech across the first year of life among infants hearing English, French, Bai, or Moroccan Arabic. Initially, we trained RAs to identify the phonation types produced by infants and used this system to code utterances in our database. To check accuracy, each utterance was independently coded by multiple RAs. Where coding differed among the RAs, we listened to the utterances together and viewed wideband spectrograms to reach agreement on which phonation type(s) occurred in these utterances. At the outset, there was high inter-rater variability, particularly in distinguishing between breathy voice and whispery voice and in coding utterances that included compound phonation types. We addressed this problem by teaching RAs to distinguish between constricted and unconstricted utterances. RAs were more likely to agree using this parameter and we began to use it as the primary criterion for coding utterances rather than individual phonation types.

While our goal was to improve inter-rater reliability, we found that this method is more effective at explaining the process of early infant phonetic development. Graphs and tables showing the incidence of individual phonation types would have obscured the most important finding: that infants from all language groups begin their first year producing primarily constricted sounds and that the incidence of constriction declines linearly over the year as infants begin to produce more unconstricted sounds. Focusing on the parameter of laryngeal constriction also showed that by the end of the first year, the incidence of constriction in infants' babbling correlates with whether or not the language they are hearing employs laryngeal constriction contrastively, e.g., more constriction persists in the babbling of infants in a Bai (tonal-register) context than in an English context [3].

The LAM also helped illustrate how infants gain articulatory control by exploring laryngeal constriction. Starting in months 3 and 4, infants begin to alternate between constricted and unconstricted phonation types within single utterances ("dynamic" utterances in our framework). These alternations in the laryngeal vocal tract help "prime" the exploration of the oral vocal tract. Before producing babbling – the most language-like of prelinguistic utterances – infants alternate between non-syllabic and syllabic sequences within the same utterance ("mixed" utterances in our framework). In these mixed utterances, infants explore coordinated use of the laryngeal and oral vocal tracts, which is essential to the emergence of speech-like utterances. The LAM was not only pedagogically more effective in training RAs, but its unique affordances made patterns in infant speech acquisition more generalizable.

3. EXAMPLE FROM A TEXTBOOK

Next, we report our experience with an introductory manual of speech sciences for Italian university students of languages and linguistics, which is currently under preparation [6]. The book is conceived of as a guide to the linguistic study of speech sounds, from production to acoustics and to hearing and perception and with practical illustrations of common procedures for the recording, archiving / FAIRification, annotation and acoustic analysis of speech. Having a modular structure, with supplementary materials for more advanced students, the book also applies laryngeal articulation ideas to phonological analysis (such as phonological contrast, neutralization, phonotactic constraint, etc.). A final section is devoted to the applications of phonetics in various professional fields (speech impairments, forensic linguistics, L2 pronunciation teaching, etc.).

Taking the LAM as the starting point for describing the dynamics of speech production introduces several novelties in the way phonetics and phonology are taught. One of them has to do with the characterization of vowel qualities. Many aspects could be discussed in this respect but we focus on only two of them for space reasons. First, although vowel quality distinctions are primarily oral, the larynx is conceived of as a generator of "inherent" quality (which most often expands over entire syllables). Second, the quadrilateral space mapping based on the traditional dichotomy between a front-back and a high-low lingual dynamics is re-conceptualized as resulting from three different articulatory dynamics [12]. Two of them, fronting and raising, are oral dimensions, controlled by the movements of the tongue and the jaw; the third one, retraction, is pharyngeal to the extent that it is controlled by the aryepiglottic sphincter. This re-conceptualization is an important departure from traditional approaches in which vowel quality distinctions are described as exclusively dependent on tongue vertical and horizontal movements (with the contribution of lip rounding, which obviously remains essential in the LAM as well).

Terminology reflects this change of perspective. For instance, for an Italian learner, [u] and [o] are raised vowels, whereas [ɔ] (like [ɑ]) is susceptible to retraction. The new conceptualization is more intuitive for learners because it is directly related to the physical dynamics of lingual muscular constriction [17]. Moreover, it nicely explains why the number of height distinctions is larger for front than for "back" vowels in natural languages. In the LAM perspective, the reason is that fronting is governed by the jaw (and the tongue), and both have more freedom to move than the laryngeal articulator, whose action unfolds (or folds) in the lower vocal tract.

The LAM has a direct impact on how the vocal tract itself is conceptually represented. A first departure from traditional approaches is that the rear of the tongue is part of the lower vocal tract. Various phenomena implying larynx constriction nicely demonstrate this. For instance, the so-called [+ATR]/[-ATR] vowel distinction that is typical of (but not limited to) Western African languages is a good example. Instead of focusing on lingual dynamics only, a LAM-based explanation is focused on the co-occurrence of tongue movements with laryngeal constriction and vertical movements (leading to pharyngeal cavity expansion or restriction). The tense/lax vowel distinction typically implemented in Germanic languages such as English and the mid-high/mid-low vowel distinction (as is typically referred to for languages such as Italian) can, consequently, be explained within a larger and much more informative picture. In such a picture, a fundamental part is played by the observation of the existing synergies between articulators. Focusing on synergies rather than on parcellization could be wrongly interpreted as a "complication" from the pedagogical point of view. Quite the contrary, it is the most realistic way of explaining how speech production works; and realism has the advantage (among various others) of raising the level of intuitiveness in the explanation of many speech phenomena.

While most examples are taken from Italian and other typologically close languages such as English or French, given the audience of the textbook, some sections are developed with data from languages spoken in regions of Asia, Africa, Oceania or the Americas. A relatively large cross-linguistic scope in exemplification is one of the consequences of using the LAM in the teaching of phonetics and phonology. This may be challenging in introductory textbooks but it is certainly highly rewarding in terms of students' awareness of phonetic diversity. For instance, the observed relationship that exists between laryngeal states (constricted vs. unconstricted larynx), phonatory qualities, and pitch variations, can be neatly exemplified by making reference to Bai tonal registers as presented in [11]. The LAM coherently accounts for how larynx constriction, tone and voice quality interact in natural languages; Bai examples are a fully-fledged exemplification of this interaction and can therefore be usefully integrated into a phonetics textbook. The supplementary materials linked with [15], available at: https://drive.google.com/drive/folders/1-Te5DhQ6NApJoDXJOePR5c_HUOBLJsec, further offer a range of video examples for teachers to use in their phonetics and phonology classes.

4. PEDAGOGY OF VOICE AND CLINICAL PHONETICS

In common parlance, voice is a generic term that can have different meanings: i) sometimes used as a synonym for speech, a confusion frequently made by patients consulting for dysphonia or dysarthria or even by doctors; ii) voice can also refer to individual characteristics of vocal behaviour, timbre or prosody; iii) to vocal categories in linguistic systems or; iv) it can refer to phonation, i.e., to the mode of production corresponding to the vibration of the vocal folds, which is a classical but very reductive definition especially in the medical field.

In the realm of pedagogy, the understanding of vocal production systems based on the LAM model is a major contribution in terms of objective description of physiology. The teaching of the complexity of the structures constituting the larynx in the lower vocal tract as an articulator of voice and speech should be part of the learning process in speech therapy and in the medical field, but also for voice pedagogues, singing teachers, voice professionals and all persons interested in vocal quality.

Clinical phonetics is the interface between experimental phonetics and the medical sciences. Synchronized multi-instrument exploration of the oral and laryngeal vocal tracts facilitates the visualization and understanding of the synergy and dynamics of speech production.

The three main objectives of including the LAM in voice and speech pedagogy are: i) to understand typical productions and atypical variations; ii) in pathology to analyze the mechanisms that cause deviant phonatory behaviours or lesions and to anticipate the consequences of surgical treatments; and iii) to optimize the therapeutic choices and rehabilitative possibilities to guide towards an efficient compensation or adaptation [15, ch. 7].

Some examples of the interdependence of the structures of the epilarynx include dysphonia related to an alteration of vocal behaviour due to misuse or overuse involving the glottic plane and having repercussions on the ventricular plane and even involving the aryepiglottic plane [9, 16]. Similarly, after laryngeal surgery [5, 7-8], when the lower part of the epilarynx is removed, the creation of a neovibrator implies the recruitment of the aryepiglottic structures and the base of the tongue, making the oral and laryngeal vocal tract inseparable. Beyond pathology, atypical productions as in certain songs [10] or vocal techniques [28] will also have recourse to the use of the epilarynx by putting in play the aryepiglottic articulator.

The extraordinary human potential for plasticity and agility of the vocal tract leads to the possibility of relearning, adaptation and compensation of the mechanisms of phonatory production. The pedagogy of the LAM contributes to an understanding of the processes of speech production and the mechanisms of its normal and pathological variability and as a guide to the best therapeutic adjustments.

5. HELPING LEARNERS CONCEPTUALLY EXPAND THE LARYNX WITH COMPUTATIONAL MODELLING

Finally, we move to the area of speech modelling. As mentioned above, the vocal folds dominate the attention of researchers and pedagogues, which has in many cases literally motivated the removal of the epilaryngeal structures. This is true both in experimental studies with excised larynges (e.g., [2]) and computational modelling studies (whether 2D lumped element, e.g., [25], or 3D finite element, e.g., [18], formulations are used).

This is not to say that epilaryngeal structures are entirely absent from such research, as they have been the focus of, or have been incorporated into, various experimental (e.g., [1, 19]) and modelling (e.g., [23]) studies. It is also not always practical to include or model every detail of a system under consideration, especially one as complex as the larynx. Removal of the epilaryngeal structures provides a clearer view of the vocal folds in studies using excised larynges, and the omission of these structures in computational modelling studies allows for savings in the cost of designing, implementing, and processing required for conducting simulations. However, from the perspective of the LAM, the absence of the epilaryngeal structures in these studies is regarded as a considerable compromise. In terms of source-filter theory, the epilaryngeal structures contribute a wealth of source generators, including vibratory action of the ventricular folds, aryepiglottic folds, epiglottis, and all of the internal vestibular mucosa spanning these structures and channels, constrictions and baffles that generate turbulent (noise) sources. The tubular spaces and side-pockets characterizing the epilaryngeal spaces provide important resonances for voice production (e.g., the "singing formant" owes its provenance to the epilarynx; see [26]) and underlies the complexity where source-filter theory breaks down (necessitating the non-linear source-filter theory outlined by Titze [27]).

Our own computational modelling work, some of which has a pedagogical orientation, endeavours to draw attention to the whole larynx, integrating the epilarynx into the design of the models as an unelidable component. The 3D-LCM (Laryngeal Constrictor Mechanism) was an early model that provided a "visual synthesis" of the articulatory and vibratory possibilities of the entire larynx [20-21].

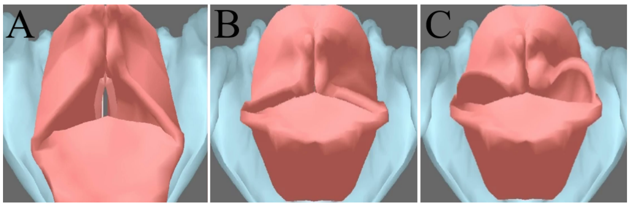

Figure 1: Three frames from the 3D-LCM model showing (A) modal phonation, (B) aryepiglotto-epiglottal stop, and (C) aryepiglottic trilling.

Fig. 1 shows three illustrative frames from this pedagogical tool, which can be manipulated to provide any desired view of the larynx (here we see the larynx as in a laryngoscopic view). Students are able to clearly visualize how the unconstricted laryngeal posture for modal phonation (A) compares to the constricted posture (B) where the aryepiglottic folds have advanced to meet the retracting epiglottis and may engage in aryepiglottic trilling (C). The instructor benefits from the ability to demonstrate the actions of the laryngeal articulator dynamically and three-dimensionally. Computer simulations can be complemented with actual physical sculpture of laryngeal structures to help students understand the anatomy in addition to its actions.

Investigative computational modelling studies of epilaryngeal vibration have been conducted that demonstrate the pulse amplitude-modulating effect that aryepiglottic vibration has on vocal fold vibration [15, ch. 3]. We have also demonstrated that the ventricular folds, through their descent and contact with the vocal folds (a manoeuvre we call vocal-ventricular fold coupling, or VVFC) – evident in ultrasound and MRI data [15] – impacts the vibratory dynamics of the vocal folds, possibly helping to yield creakiness in phonation and aiding in glottal stop production. Finally, we have explored 3D finite-element modelling as a tool to discover the degree of biomechanical quantality (in the sense of Stevens [24]) that exists within the larynx thanks in large part to the presence of the epilaryngeal structures [22]. Apart from video demonstrations, these models can be used in the classroom in an interactive fashion similar to the 3D-LCM, including dynamic simulation if the computational demands of the model setup are manageable enough to be run at a reasonable frame rate.

Because models have a visual appeal and provide a simplified means of understanding the behaviour of complex structures and systems, it is important for both laryngeal research and pedagogy that every effort be made to acknowledge and incorporate the epilaryngeal structures. Doing so will help us to move beyond the conceptualization of the larynx as only a pair of vocal folds.

6. RECAPITULATION

This paper has provided specific examples of some of the pedagogical benefits of the Laryngeal Articulator Model. We hope the readers find these useful in their own teaching practices and that their students will benefit from the supplementary video materials accompanying [15], which are freely available at: https://drive.google.com/drive/folders/1-Te5DhQ6NApJoDXJOePR5c_HUOBLJsec

7. REFERENCES

[1] Alipour, F. and Finnegan, E.M. 2013. On the acoustic effects of the supraglottic structures in excised larynges. *Journal of the Acoustical Society of America* 133, 5, 2984–2992.

[2] Alipour, F., Finnegan, E.M. and Jaiswal, S. 2013. Phonatory characteristics of the excised human larynx in comparison to other species. *Journal of Voice* 27, 4, 441–447.

[3] Benner, A. 2009. *Production and Perception of Laryngeal Constriction in the Early Vocalizations of Bai and English Infants*. PhD dissertation, University of Victoria.

[4] Benner, A., Grenon, I. and Esling, J.H. 2007. Infants' phonetic acquisition of voice quality parameters in the first year of life. *Proc. 16th ICPhS* Saarbrücken, 2073–2076.

[5] Biacabe, B., Crevier-Buchman, L., Laccourreye, O., Hans, S. and Brasnu, D. 2001. Phonatory mechanisms after vertical partial laryngectomy with glottic reconstruction by false vocal fold flap. *Annals of Otology, Rhinology & Laryngology* 110, 935–940.

[6] Calamai, S. and Celata, C. In prep. *Manuale di scienze del parlato*. Carocci.

[7] Crevier-Buchman, L., Laccourreye, O., Weinstein, G., Garcia, D., Jouffre, V. and Brasnu, D. 1995. Evolution of speech and voice following supracricoid partial laryngectomy. *Journal of Laryngology & Otology* 109, 410–413.

[8] Crevier-Buchman, L., Maeda, S., Brasnu, D. and Vaissière, J. 2003. Perceptual and acoustic correlation in consonant identification after partial laryngectomy. *Proc. 15th ICPhS* Barcelona, 2361–2364.

[9] Crevier-Buchman, L., Mattei, A. and Giovanni, A. 2019. Forçage vocal. *EM Consulte Oto-Rhino-Laryngologie*, 20-752-B-10.

[10] Crevier-Buchman, L., Pillot-Loiseau, C., Rialland, A., Vincent Narantuya, C. and Desjacques, A. 2012. Analogy between laryngeal gesture in Mongolian Long Song and supracricoid partial laryngectomy. *Clinical Linguistics and Phonetics* 26, 86–99.

[11] Edmondson, J., Esling, J.H. and Li, S. 2021. Jianchuan Bai. *Journal of the International Phonetic Association* 51, 3, 490–501.

[12] Esling, J.H. 2005. There are no back vowels: The Laryngeal Articulator Model. *The Canadian Journal of Linguistics / La revue canadienne de linguistique,* 50, 13–44.

[13] Esling, J.H. 2009. The control of laryngeal constriction and the emergence of speech in infants in the first year of life. In: Fant, G., Fujisaki, H. and Shen, J. (eds), *Frontiers in Phonetics and Speech Science: Festschrift for Prof. Wu Zongji's 100th Birthday*. The Commercial Press, 191–203.

[14] Esling, J.H. 2012. The articulatory function of the larynx and the origins of speech. *Proc. Annual Meeting Berkeley Linguistics Society* 38, Berkeley (LSA eLanguage).

[15] Esling, J.H., Moisik, S.R., Benner, A. and Crevier-Buchman, L. 2019. *Voice Quality: The Laryngeal Articulator Model*. Cambridge University Press.

[16] Esling, J.H., Zeroual, C. and Crevier-Buchman, L. 2007. A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. *Proc. 16th ICPhS* Saarbrücken, 585–588.

[17] Honda, K. 1996. Organization of tongue articulation for vowels. *Journal of Phonetics* 24, 39–52.

[18] Hunter, E.J., Titze, I.R. and Alipour, F. 2004. A three-dimensional model of vocal fold abduction/adduction. *Journal of the Acoustical Society of America* 115, 4, 1747–1759.

[19] Kimura, M., Sakakibara, K.-I., Imagawa, H., Chan, R., Niimi, S. and Tayama, N. 2002. Histological investigation of the supra-glottal structures in human for understanding abnormal phonation. *Journal of the Acoustical Society of America* 112, 2446.

[20] Moisik, S.R. 2008. *A Three-Dimensional Model of the Larynx and the Laryngeal Constrictor Mechanism: Visually Synthesizing Pharyngeal and Epiglottal Articulations Observed in Laryngoscopy*. Masters thesis, University of Victoria.

[21] Moisik, S.R. and Esling, J.H. 2007. 3D auditory-articulatory modeling of the laryngeal constrictor mechanism. *Proc. 16th ICPhS* Saarbrücken, 373–378.

[22] Moisik, S.R. and Gick, B. 2017. The quantal larynx: The stable regions of laryngeal biomechanics and implications for speech production. *Journal of Speech Language and Hearing Research* 60, 3, 540–560.

[23] Sakakibara, K.-I., Imagawa, H., Konishi, T., Kondo, K., Murano, E.Z., Kumada, M. and Niimi, S. 2001. Vocal fold and false vocal fold vibrations in throat singing and synthesis of khöömei. *Proc. International Computer Music Conference* La Habana, 135–138.

[24] Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3–45.

[25] Story, B.H. and Titze, I.R. 1995. Voice simulation with a body-cover model of the vocal folds. *Journal of the Acoustical Society of America* 97, 2, 1249–1260.

[26] Sundberg, J. 1974. Articulatory interpretation of the 'singing formant'. *Journal of the Acoustical Society of America* 55, 4, 838–844.

[27] Titze, I.R. 2008. Nonlinear source–filter coupling in phonation: Theory. *Journal of the Acoustical Society of America* 123, 5, 2733.

[28] De Torcy, T., Clouet, A., Pillot-Loiseau, C., Vaissiere, J., Brasnu, D. and Crevier-Buchman, L. 2014. A videofiberscopic study of laryngopharyngeal behaviour in the human beatbox. *Logopedics Phoniatrics Vocology* 39, 1, 38–48.