Articulatory model of the epiglottis

Yves Laprie, Benjamin Elie, Pierre-André Vuissoz, Anastasiia Tsukanova

Articulatory model of the epiglottis

Yves Laprie[1], Benjamin Elie[1], Pierre-André Vuissoz[2] and Anastasiia Tsukanova[1]

[1]*CNRS/INRIA/UL LORIA,* [2]*UL/INSERM U947 IADI, Nancy, France*

Yves.Laprie@loria.fr

Abstract

This works describes the construction of the articulatory model of the epiglottis from MRI images. This new model takes into account the influences of the mandible, tongue and larynx via a multi-linear regression applied to the contours of the epiglottis. Once these influences are removed from the contours principle component analysis is applied to the control points of the B-spline representing the centerline of the epiglottis.

Keywords: articulatory model, epiglottis

1. Introduction

The position of the epiglottis has an acoustic influence in the lower part of the vocal tract by decreasing the transverse area when it is contacting the tongue root, or when it reduces the volume of the laryngeal vestibule. Conversely, it is assumed to have a negligible acoustic impact when the pharyngeal cavity is wider (as /i/ for instance) and the epiglottis is well detached from the tongue. Epiglottis is a cartilage connected to the hyoid bone via the thyroid cartilage. Unlike tongue and lips whose intrinsic muscles are at the origin of a substantial part of their deformations, those of the epiglottis are mainly due to the influence of other articulators, i.e. the tongue, larynx, and indirectly mandible. A global articulatory model of the vocal tract should take these properties into account.

2. Description of the model

Similarly to the model of Maeda Maeda 1979 the deformation modes of our articulatory model (Y. Laprie and Busset 2011; Yves Laprie, Elie, and Tsukanova 2015) are obtained by applying Principal Component Analysis (PCA) on the articulator contours (mandible, tongue, lips…) by using prior knowledge of the interactions between articulators. For instance, the tongue is attached to the mandible whose movement is subtracted from the tongue contours before applying PCA. So far our epiglottis articulatory model only incorporated the indirect influence of the mandible and the influence of the tongue via a collision model. Its deformation modes thus integrated movements due to the tongue and larynx.

Other deformation modes were due to the errors of delineation. In the case of the tongue, these errors are marginal, or at least give rise to deformation modes coming after the genuine deformations whose amplitude is bigger. On the other hand, the width of epiglottis is small on the image, and the errors of delineation, whether manual or automatic, are of the same order of magnitude as genuine deformations. Consequently, a PCA applied without precaution will mix both types of deformation.

To prevent the apparition of these spurious deformation components the epiglottis was approximated as a thick curve, and only the centerline of the epiglottis was analyzed. As a matter of fact, the centerline was determined after delineation of all the epiglottis contours, and the width was set as the average width of all these contours in the upper part where the two epiglottis edges are clearly visible (see Fig. 1). The height of the upper part (where both contours are visible) is adjusted by hand to fit the contours extracted from images. The centerline is approximated as a B-spline and represented by its control points P_l ($0 \leq l < M$ where M is the number of control points) in the form of a two-coordinate vector, and the reconstruction of the epiglottis from the centerline amounts to draw a line at a distance of half the width from the centerline.

In order to take into account the influences of the jaw, larynx and tongue, we applied multiple linear regression Rao, R. Toutenburg, and Heumann 2008 on the control points P_l ($0 \leq l < M$) of the epiglottis contour:

$$P_l = jaw_{0,l} B_0 + \sum_{j=0}^{j=T-1} tg_{j,l} C_j + lx_{0,l} D_0 + E_l$$

where $jaw_{0,l}$ is the control factor of the first linear component of the jaw in the global articulatory model and B_0 the two-coordinate regression vector for the jaw, $tg_{j,l}$ the control factors of the T first linear components of the tongue and C_j the corresponding regression vector, $lx_{0,l}$ the first control factor of the larynx component and D_0 the corresponding regression vector, and E_l is the residue vector.

Actually, since the centerline is represented by the M control points, each centerline occurrence is represented by a $2 \times M$ vector formed by the coordinates of the M points. Each of the N occurrences of the epiglottis is represented by a $2 \times M$ vector named P_i with $0 \leq i < N$. These contour vectors P_i are the observations, and the control factors of the jaw $jaw_{0,i}$, tongue $tg_{j,i}$ and larynx $lx_{0,i}$ are the input explanatory variables. For the jaw and larynx we kept only the first factor (jaw_0, resp. lx_0), which is in both cases the most informative one. For the tongue, which is in front of the epiglottis, we kept the first six deformation factors $tg_{0..5}$. All the variables are centralized and therefore the intercept can be ignored.

By grouping all the explanatory factors jaw_0, tg_j, and lx_0, in a vector of $K = T + 2$ coordinates which are named X_j, the previous equation is

$$P_i = \sum_{j=0}^{j=K-1} b_j X_i + e_i$$

or in a matrix form

$$P = XB + E$$

, where P is the $N \times 2M$ matrix of the observations, i.e. the control points of the centerlines, X is the $N \times K$ matrix of the explaining articulatory factors, B is the $K \times 2M$ regression matrix and E is the $N \times 2M$ matrix of residue not explained by jaw, tongue and larynx. B is given by

$$B = (X^t X)^{-1} X^t Y$$

.

Then, the contributions of the jaw, tongue and larynx can be subtracted from the observations of the epiglottis, i.e. $P - XB$, and PCA can be applied to the residue E. Given the nature of epiglottis, i.e. a cartilage, the number of relevant linear components should be small.

We applied this analysis scheme to two sets of data: midsagittal slice of 3D MRI images recorded at IADI of 90 static syllables covering vowels and blocked CV articulations made by a male speaker and a corpus of 1021 mid-sagittal images from an X-ray film (from the DOCVACIM database) made of short sentences recorded by a female speaker. Tab. 1 shows that

σ in mm	male (static)	female (dynamic)
total	8.44	14.32
regression	5.08	6.51
PCA 1st	2.66	3.17
PCA 2nd	1.33	1.73
PCA 3rd	0.78	0.93

Table 1: Standard deviations in millimeter of the epiglottis: total, after subtracting the influence of other articulators, after subtracting 1^{st}, 2^{nd} and 3^{rd} linear components.

most of the variance is explained by the articulators influencing the epiglottis and that two linear components are enough to approximate the epiglottis fairly accurately.

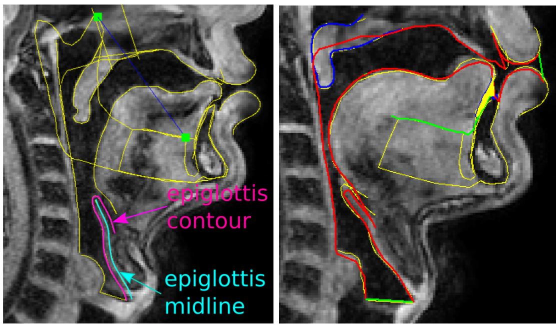

Figure 1: The epiglottis contour and the centerline (left), and the complete articulatory model in red approximating the vocal tract for the articulation /sø/ (right)

3. Concluding remarks

The regression stage enables most of the extrinsic influences exerted on the epiglottis to be removed before PCA. It should be noted that these influences explain as much variability as the intrinsic components. This new model is currently used with our articulatory copy synthesis approach.

Fig.2 and 3 show that the first factor corresponds to a vertical movement, although the second corresponds to a horizontal movement in the lower and mid part of the epiglottis which contributes to the constriction of the whole laryngeal vestibule. This retraction is not explicitly used in French which is the

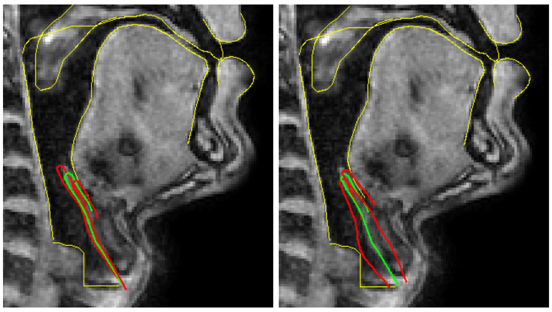

Figure 2: First two deformation modes of the epiglottis (male speaker). The green line is the neutral position (middle)

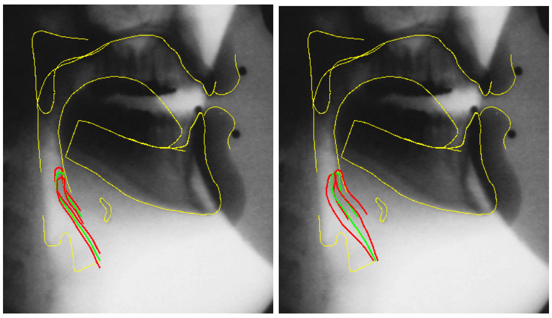

Figure 3: First two deformation modes of the epiglottis (female speaker). The green line is the neutral position (middle)

mother tongue of the two speakers but plays a determining role in the production of glottal consonants in Arabic or Hebrew Laufer and Condax 1981; Esling, Fraser, and Harris 2005. We plan to record speakers of these two languages to investigate this question.

4. Acknowledgements

This work benefited from the support of the project ArtSpeech of ANR (Agence Nationale de la Recherche).

5. References

Esling, J. H., K. E. Fraser, and J. G. Harris (2005). "Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka)". In: *Journal of Phonetics* 33.4, pp. 383–410.

Laprie, Y. and J. Busset (2011). "Construction and evaluation of an articulatory model of the vocal tract". In: *19th European Signal Processing Conference - EUSIPCO-2011*. Barcelona, Spain.

Laprie, Yves, Benjamin Elie, and Anastasiia Tsukanova (2015). "2D Articulatory Velum Modeling Applied to Copy Synthesis of Sentences Containing Nasal Phonemes". In: *International Congress of Phonetic Sciences*. Glasgow, United Kingdom. URL: `https://hal.inria.fr/hal-01188738`.

Laufer, A. and I. D. Condax (1981). "The function of the epiglottis in speech". In: *Language ad Speech* 21.1, pp. 39–62.

Maeda, S. (1979). "Un modèle articulatoire de la langue avec des composantes linéaires". In: *Actes 10èmes Journées d'Etude sur la Parole*. Grenoble, pp. 152–162.

Rao, C. R., H. Shalabh andC. R. Toutenburg, and C. Heumann (2008). *Linear Models and Generalizations*. Springer.