

A three-dimensional linear articulatory model of velum based on MRI data

Antoine Serrurier & Pierre Badin

Institut de la Communication Parlée
UMR 5009, CNRS, Université Stendhal, INPG, Grenoble, France
{Antoine.Serrurier, Pierre.Badin}@icp.inpg.fr

Abstract

In the framework of studies on nasality, we have attempted to develop a 3-D articulatory model of velum. Sets of 25 sagittal MRI images have been collected for one French subject sustaining 46 articulations. Contours of the velum have been manually drawn from these images; a common 3-D triangular mesh has then been fitted to these contours in order to constitute 3-D representations of the velum outline. This resulted in 46 meshes of 5239 vertices defined by their 3-D coordinates.

In order to develop a velum model free of the influence of the tongue, the analysis was first limited to the 28 articulations where no tongue – velum contact occurs: it was found that the first component extracted from these data by Principal Component Analysis can explain as much as 79 % of the variance of the complete 3-D shape of the velum, while the second component was meaningless in terms of articulatory movements. It was also found that the same model could explain 71 % of the variance for the corpus of the 18 articulations where tongue – velum contact occurs. It was finally shown that the vertical coordinate of a single flesh point attached to the downward face of the velum is enough to predict the complete 3-D shape of the free velum.

1. Introduction

The problem of nasality is complex and has given rise to a large number of studies, from both perception and production points of view (for example [1], [2], [3]). The nasality feature is related to the velum position: lowering the velum, and thus opening the velopharyngeal port, is a simple gesture that induces strong and complex changes in the vocal tract acoustical behaviour. The realisation of nasality involves (1) an articulatory level that deals with the shape of the articulators and their articulatory degrees of freedom, and (2) a control level that deals with the coordination of these articulators. The present article describes the two- and three-dimensional articulatory models of velum without mechanical perturbations from the tongue based on one specific subject and following the process described by the authors in [4].

2. Modelling approach

2.1. A subject oriented linear modelling approach

Following a method already proven for orofacial articulatory modelling in [5], the 3-D geometry of the velum will be modelled as the weighted sum of a small number of linear components. These components will be extracted by linear analysis from a set of vocal and nasal tract shapes representative of the speech production capabilities of the subject. The analysis is based on both *Principal Component Analysis (PCA)* and *linear regression*. The weights of the sum

constitute the *articulatory control parameters* associated with the components: a given set of values of these parameters produces a given single shape of the organs.

Two stages are necessary to develop the model: the construction of the database of shapes from images (described in section 3), and then the analysis of the corresponding 3-D coordinates (detailed in section 4)

Note that our approach is *subject-oriented*, i.e. that the model will be based on a single (French) subject. This avoids to merge the physiological characteristics and control strategy of different subjects, which can be quite different. This approach can be naturally criticized as it is speaker dependent and may not generalize extensively. However, in this first approach, the model does not intend to cover all the possible aspects of nasal articulation, but to explain a possible mechanism of articulatory movements.

2.2. The corpus

The corpus consisted of a set of artificially sustained articulations designed as to cover the maximal range of articulations: the French oral and nasal vowels [a e i y u o ø ɔ œ ã ě œ ɔ̃], and the consonants [p t k f s ʃ m n ɳ l] in three symmetrical contexts [a i u]. This corpus was supplemented by two specific articulations frequent in speech: the *rest* articulation (in a rest position) and the *prephonatory* articulation (in the preparatory phase preceding phonation). Finally we have a corpus of 46 French phonemes.

As described above, the corpus contains only 46 static articulations with many more oral articulations than nasal ones. The use of a limited number of target articulations to build a model is justified by the investigation of Badin et al. [6] who showed that a limited number of target articulations leads to an accurate articulatory model. The corpus intends to cover the maximal range of the articulators positions, and the lack of balance between orals and nasals will not change the ability of the model to reach all possible articulations.

3. Data acquisition and processing

In order to be able to develop an articulatory model following the approach described above, it is needed to obtain a 3-D surface representation of the velum and of the rigid structures for each articulation of the corpus, based on stacks of CT and MR images.

3.1. Data acquisition

A Computer Tomography scan of the head of the subject was made, to serve as a reference. A stack of 149 axial images with a size of 512×512 pixels, a resolution of 20 pixels/cm, and an inter slice space of 0.13 cm, spanning from the neck to the top of the head, was recorded for the subject at rest (see one example image in Fig. 1a). These images allow making the distinction between bones, soft tissues and air, but do not allow discriminating different soft tissues. They will be used

to locate bony structures and to determine accurately their shapes for reference (see section 3.2).

Stacks of sagittal MR images were recorded for the French subject sustaining artificially during about 45 sec. each of the 46 articulations of the corpus (Fig. 1b illustrates vowel [a]). The subject was instructed to artificially sustain the articulation throughout the whole acquisition time. The consonants were produced in three different symmetrical vocal contexts [VCV], V belonging to [a i u]. A set of 25 sagittal images with a size of 256×256 pixels, a resolution of 10 pixels/cm and an inter slice distance of 0.4 cm was obtained for each articulation. These images allow to make the distinction between soft tissues and air, and to discriminate soft tissues, but not to distinguish clearly the bones. Note that for both imaging techniques, the subject was in a supine position, which may alter somehow the natural shape of articulators.

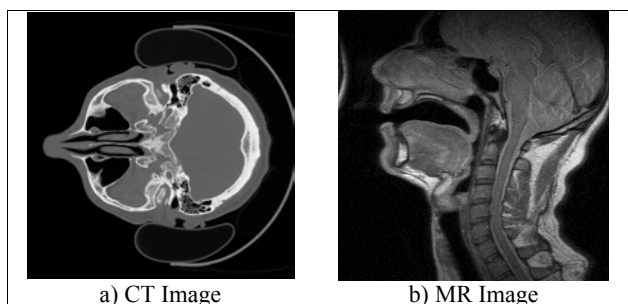


Figure 1: Example of original axial CT image (a) and sagittal MR image for [a] articulation (b)

Due to the complexity of the contours of the various organs, to the relatively low resolution of the images, and to the need of an accurate reconstruction of the organs, the extraction of contours has been performed manually, plane by plane. In order to extract maximal information from these images during the manual edition, they have been supplemented by extra sets of images reconstructed by intersection of the initial stacks with planes having a different orientation.

For the MR images, the initial sagittal stack was resliced in images perpendicular to the vocal tract, considering that they will be used to extract organs shapes around the vocal tract (e.g. velum, tongue, etc.). They were thus resliced for each articulation in 27 planes orthogonal to the midsagittal plane and intersecting it along a semipolar grid, as illustrated in Fig. 2a. Each new image is arbitrarily given a size of 200×100 pixels and a resolution of 10 pixels/cm. Finally we dispose of two redundant stacks of MR images for each articulation.

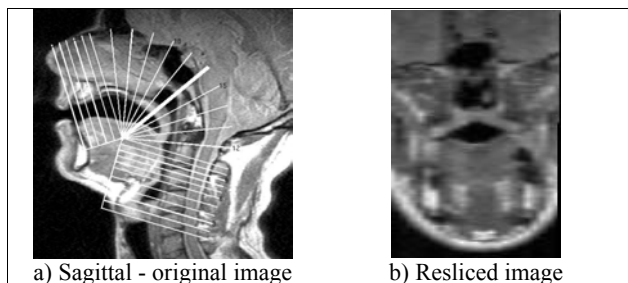


Figure 2: Example of MR images reslicing: one perpendicular image for an [a] articulation

3.2. From images to 3-D shapes

According to the process of extraction of the 3-D shapes from CT and MR images described in [4], 6 stages are necessary for the velum shapes extraction: (1) manual edition of each rigid organs (jaw, hyoid bone, hard palate, nasal passages, nostrils and various paranasal sinuses), plane by plane (see Fig. 3a for the jaw); (2) expansion of the set of all 2-D plane contours into a 3-D coordinate system (see Fig. 3c for the jaw) and processing through a 3-D meshing reconstruction software service provided by the Geometrica Research Group at INRIA ("<http://cgal.inria.fr/Reconstruction/index.html>") to form a 3-D surface meshing based on triangles (Fig. 3e); (3) projection of the intersections of the 3-D rigid surfaces extracted above with the planes corresponding to the MR images on these images, in order to provide some useful anchor points for the interpretation of the images and for the tracing of the soft structure contours (part of the dotted-lines of the Fig. 3b); (4) manual edition of the velum (solid line of the Fig. 3b for the [a] articulation) plane by plane in the 46 stacks of MR images on which were superposed previously intersected rigid organs shapes in order to provide reference information of rigid organs on the MR images; (5) expansion of the set of all 2-D plane contours of the velum into the same a 3-D coordinate system (see Fig. 3d for the velum in the [a] articulation); (6) fitting of a generic mesh of velum (defined arbitrarily as the mesh of the velum extracted through the current process for the [ã] articulation; see Fig. 3f) on each 3-D shape obtained at step 5 for the 46 articulations through the matching software *TestRigid* developed at the TIMC laboratory in Grenoble [7] in order to have a same number of points for the velum for the 46 articulations.

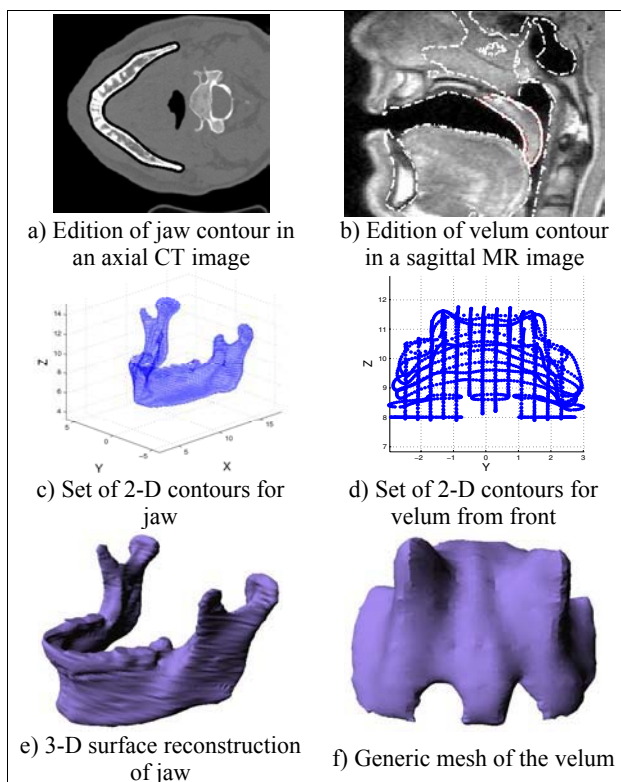


Figure 3: Illustration of 3-D surface reconstruction process from images

This reconstruction process finally provides a set of 46 velum surfaces described as a triangular mesh with 5239 vertices defined by their 3-D coordinate in a common reference coordinate system. The matching process of the generic mesh to the MRI data has resulted in a root mean square (RMS) reconstruction error of 0.06 cm. This set of data constitutes the basis for the articulatory modelling of the velum detailed in the following.

4. Articulatory model of the velum

It has been observed that the 46 articulations of the corpus can be divided in two groups: (1) one group of 18 articulations for which the velum can be considered in mechanical contact with the tongue in the midsagittal plane (a minimal distance between velum and tongue under the threshold of 0.075 cm): [a u o ɔ ã ã m^u s^a s^u ʃ^a p^u k^a k^u ɸ^a ɸ^u rest prephonation], and (2) another group of 28 articulations for which the velum can be considered free of contact, and thus of mechanical perturbation by the tongue : [ɛ e i y ø œ ã m^a mⁱ n^a nⁱ n^u f^a fⁱ f^u sⁱ ʃⁱ ʃ^u p^a pⁱ t^a tⁱ t^u kⁱ l^a lⁱ l^u]. The present section proposes a two- and a three-dimensional linear models of velum for movements without mechanical perturbation from the tongue.

4.1. Two -dimensional model

A representation of the velum in the midsagittal plane has been obtained by determining the intersection of the 3-D mesh with the midsagittal plane, and by resampling the contour obtained by a fixed number of 202 2-D points for each of the 28 articulations without tongue contact. A *direct PCA* was applied to these 202×2 variables in order to extract a few articulatory control parameters by exploiting the correlations between neighbouring points and due to the physical continuity of the organ. Table 1 shows that almost 93 % of the variance of all the velum points is explained by the first PCA parameter, *VL_2D*, and that the associated RMS reconstruction error is 0.07 cm. The extra 3 % variance explained by the second PCA parameter, *P2_2D*, does not correspond to meaningful articulatory movements, and thus *P2_2D* has not been considered for the model.

The effect of the *VL_2D* parameter on the velum is illustrated in Fig.4 (a & c) for two extreme values (-2 and +2). The action of this parameter corresponds to a simultaneous motion of the velum along the vertical and horizontal directions. Considering its orientation and its prime importance for speech [8], the *levator veli palatini* muscle can be thought to be much involved in this movement.

4.2. Three-dimensional model

The same approach was applied to the set of 3-D velum shapes on the same corpus: a direct PCA was applied to the 28 observations of the 5239×3 variables. The results are displayed in Table 2. The first PCA parameter *VL_3D* explains about 79 % of the full 3-D variance of all the points of the velum. The difference of about 14 % of explained variance between the 2-D and 3-D models by the first parameter is mainly associated to the difference in the number of variables (15717 for 3-D vs. 404 for 2-D) and the associated noise. However, the RMS reconstruction error with only the first parameter are very similar (compare Tables 1 & 2). For the same reasons as for the 2-D case, the second parameter *P2_3D* is not taken into account for the model.

The effects of the control parameter *VL_3D* on the 3-D velum shape is illustrated in Fig. 4 (b & d) for two extreme values (-2 and +2). It appears clearly that the movements in 2-D and 3-D are comparable, in term of range and direction of deformation.

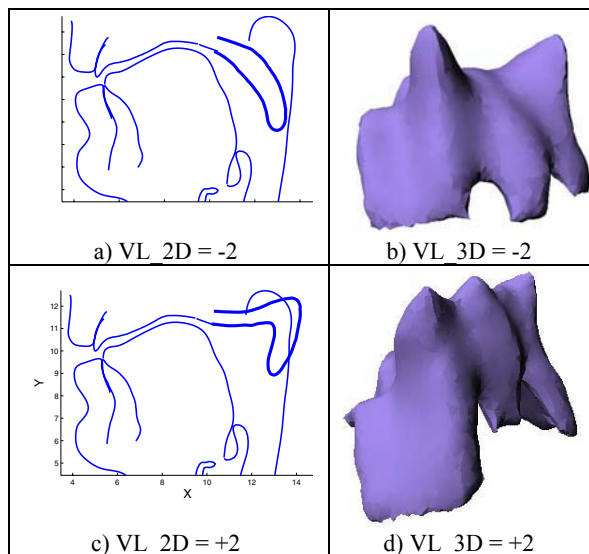


Figure 4: Representation of the velum in 2-D (a & c) and 3-D (b & d) for opposite articulatory control parameters values for the first components obtained by PCA on 2-D and 3-D data.

4.3. Relations between 3-D shapes and midsagittal contours

As a similar behaviour was observed for the 2-D and 3-D models (one principal parameter, same range of deformation, same direction of deformation, same RMS error reconstruction), we analysed the relations between the midsagittal contours of the velum and the full 3-D shapes. A high correlation of about 0.99 was found between the two principal PCA parameters *VL_2D* and *VL_3D* extracted from 2-D and 3-D data. This correlation is confirmed by the very similar explanation of full 3-D variance of the velum by the two parameters, as seen in Table 2. We can conclude that the whole 3-D velum shape can be predicted from its contour in the midsagittal plane (interestingly, Badin *et al.* [5] found similar results for the tongue and for the lips and face). Rossato *et al.* [1] have recorded the midsagittal coordinates of a coil attached to the velum by means of an ElectroMagnetic midsagittal Articulograph for the same subject over a large corpus of phoneme sequences and sentences. As the 3-D shape of the velum is primarily controlled by one parameter, we wondered if the coordinates of this coil could be used to predict the velum shape for this corpus. Therefore, we determined the point on the 2-D velum contour that is the closest to the location of the coil for all the 28 articulations of our MRI corpus. We found that the vertical coordinate of this point, *velum height* (VH), can predict the 2-D and 3-D velum shape with very small additional errors in comparison to *VL_2D* and *VL_3D* parameters (see Tables 1 and 2). This confirms that it will be possible to drive the 3-D velum shape from this recordings.

Table 1: Explained variance and RMS reconstruction error for the 28 2-D velum contours by the direct PCA parameters and by the velum height

Param.	Explained variance 2D (%)	Cumul. ex. variance 2D (%)	RMS error (cm)
VL_2D	92.89	92.89	0.07
P2_2D	3.05	95.94	0.06
VH	91.81		0.08

Table 2: Explained variance and RMS reconstruction error for the 28 3-D velum shapes by the direct PCA parameters, the 2-D PCA parameters and by the velum height

Param.	Explained variance 3D (%)	Cumul. ex. variance 3D (%)	RMS error (cm)
VL_3D	79.13	79.13	0.08
P2_3D	7.17	86.3	0.06
VL_2D	78.29		0.08
VH	77.17		0.08

4.4. Perturbations of the velum by the tongue

The models described above do not include the articulations where the tongue is in contact with the velum. We suspect that a part of the velum shape for these articulations may be simply due to the backward movement of the tongue that will push the velum backwards and possibly upwards. In order to quantify the mechanical perturbation brought by the tongue, we have determined the parameters VL_2D, VL_3D and VH for the 18 articulations with velum – tongue contact, by inverting the direct model (using the pseudo inverse of the matrix of the linear direct model). We have then used these control parameters to predict the 2-D or 3-D shape of the velum. We observed that the variance of the data (2-D and 3-D) explained by these parameters for the 18 articulations is about 10 % lower than that for the 28 articulations without contact (compare Tables 1, 2 & 3), while the RMS reconstruction error is higher by about 0.05 cm for the 2-D midsagittal contours and by about 0.01 cm for the 3-D shapes. These observations highlight the influence of the tongue on the velum shape, particularly in the midsagittal plane (probably due to a contact tongue – uvula). A further investigation must be lead on the influence of the tongue on the velum in order to improve the model and to take into account a contact effect, possibly by introducing non-linearity.

Table 3: Explained variance and RMS reconstruction error for the 18 2-D velum contours and 3-D velum shapes by VL_2D, VL_3D and VH.

Param.	Explained variance (%)	RMS error (cm)
VL_2D	81.07 (2D)	0.13 (2D)
VH	81.04 (2D)	0.13 (2D)
VL_2D	69.05 (3D)	0.09 (3D)
VL_3D	70.65 (3D)	0.09 (3D)
VH	69.10 (3D)	0.09 (3D)

5. Conclusion

We have shown that the three-dimensional description of the velum shape for a set of 28 sustained articulations where no contact occurs between tongue and velum can be modelled by means of a linear model with one control parameter. This control parameter, that can be interpreted as the height of a flesh point attached to the downward face of the velum in the midsagittal plane, can predict both 2-D and 3-D shapes with total variance explanations of respectively 93 and 79 %, and RMS reconstruction errors of 0.07 and 0.08 cm. Interestingly, we have found that the whole 3-D shape of the velum can be predicted from its midsagittal contour.

We intend now to take into account – possibly in a non-linear way – the perturbations introduced by the contact with the tongue to build a complete velum model. A longer-term objective is to build a complete vocal tract articulatory model that could provide an accurate basis for vocal tract acoustical modelling.

6. Acknowledgements

The medical images have been acquired at the Radiology Department of the Grenoble Regional University Hospital, in collaboration with the research unit INSERM/UJF 594 (Christoph Segebarth). We acknowledge the help from the GMCAO team at TIMC for the use of the matching software *TestRigid* (Yohan Payan, Franz Chouly, Maxime Béar), and from Andreas Fabri at the INRIA-Geometrica group for the surface reconstruction software.

7. References

- [1] Rossato, S., Badin, P. & Bouaouni, F. (2003) "Velar movements in French: an articulatory and acoustical analysis of coarticulation". In *Proceedings of the 15th ICPHs*, Barcelona, Spain: 3141-3144.
- [2] Teixeira, A., Vaz, F. & Príncipe, J.C. (2000) "Nasal vowels following a nasal consonant". In *Proceedings of the 5th SSP*, Kloster Seeon, Germany: 285-288
- [3] Huffman, M. K. & Krakow, R. A. (1993). *Phonetics and phonology. Nasals, Nazalisation and the Velum. Vol 5*. Academic Press Inc
- [4] Serrurier, A. & Badin, P (2005). "Towards a 3D articulatory model of velum based on MRI and CT images", *ZAS Papers in Linguistics* (in press).
- [5] Badin, P., Bailly, G., Revêret, L., Baciú, M., Segebarth, C. & Savariaux, C. (2002) "Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images". *Journal of Phonetics*, 30(3): 533-553
- [6] Badin, P., Bailly, G., Raybaudi, M. & Segebarth, C. (1998). A three-dimensional linear articulatory model based on MRI data. In *Proceedings of the Third ESCA / COCOSDA International Workshop on Speech Synthesis*, Jenolan Caves, Australia, December 1998, 249-254
- [7] Couteau, B., Payan, Y. & Lavallée, S. (2000). "The mesh-matching algorithm: an automatic 3D mesh generator for finite element structures". *Journal of Biomechanics*, 33: 1005-1009.
- [8] Bell-Berti, F. (1993). "Understanding velic motor control: studies of segmental context". In Huffman & Krakow (Eds). *Phonetics and phonology. Nasals, Nazalisation and the Velum. Vol 5. Academic Press Inc*, 63-85