[Invited Paper]

# Three-dimensional modeling of speech organs: Articulatory data and models

Pierre BADIN & Antoine SERRURIER

Speech Communication Institute, UMR 5009
CNRS – INPG – Université Stendhal, Grenoble, France

46 avenue Félix Viallet, 38031 Grenoble Cedex 01, France

E-mail: (Pierre.Badin, Antoine.Serrurier)@icp.inpg.fr

**Abstract** Volume images of tongue, jaw, velum, nasopharyngeal wall, etc., were acquired by MRI on one French subject uttering a corpus of sustained articulations. Supplementary images of jaw, hard palate, and nasal cavities were acquired by CT. The three-dimensional outlines of these organs are represented by the vertices of triangular surface meshes. Using linear component analysis, the tongue was found to possess five degrees of freedom, including one for jaw movement, and the velum and nasopharyngeal port two degrees of freedom. These parameters can be interpreted in phonetic / biomechanical terms control a linear articulatory model of speech production.

**Key words** Speech production, articulatory modeling, MRI, linear analysis, PCA.

## 1. Introduction

For a very long time, articulatory modeling of vocal tract and speech production organs has been essentially limited to the midsagittal plane. But progress and refinements brought into this domain led to a point where three-dimensional (3D) modeling has become unavoidable. The first motivation for such an approach is that the area function needed for calculating the sounds produced by an articulation specified in the midsagittal plane has to be inferred from the sole associated midsagittal contours, problem which is obviously impossible to solve as long as no other information is available on the transverse shape and size of the vocal tract. In particular, there are articulatory situations that can lead to occlusions in the midsagittal plane, while the vocal tract is not actually occluded on each side of this plane: (i) lateral consonants that are characterized by the presence of a complete apico-alveolar closure in the midsagittal plane while lateral channels are maintained open; (ii) nasal vowels, for which the uvula – an appendix of the velum in the midsagittal region – is often in contact with both the back of the tongue and the nasopharyngeal wall, creating an occlusion in the midsagittal plane, while lateral channels remain open. The detailed 3D knowledge of the vocal tract shape is important to deal in a more realistic way with the aerodynamic models needed for speech production studies. Finally, a 3D display of articulators can be useful in the domain of speech rehabilitation.

We had conducted a first 3D modeling study (Badin, Bailly, Revéret, Baciu, Segebarth & Savariaux (2002)) using a set of three complementary stacks of MRI images (one axial stack in the laryngo-pharyngeal region, one oblique stack in the velar region, and one coronal stack in the front region including lips). We experienced difficulties in determining tongue tip, lips or velar region with satisfactory accuracy. It was thus decided to develop similar models with much better defined tongue tip, sublingual cavity, lateral cheek cavities, velum and lips, based on sets of 25 sagittal MRI images that allow a better accuracy in a number of situations and a more reliable reconstruction of transverse images.

The present article reports our attempts to reconstruct 3D jaw, tongue, velum and nasopharyngeal wall positions and shapes from MRI and CT data for a single subject uttering a more comprehensive corpus of sustained articulations in French, and to develop corresponding 3D linear articulatory models.

## 2. A subject-oriented linear modeling approach

Our modeling approach is described in details in Badin *et al.* (2002). In the framework of *speech robotics*, the speech apparatus is viewed as a *plant* (an articulatory model) driven by a *controller* so as to recruit articulators and coordinate their movements in order to generate audio-visual speech. This concept implies the notion of a relatively small number of *degrees of freedom* (henceforth DoF) for the articulatory plant, *i.e.* the specification, for each articulator, of the limited set of movements that it can execute independently of the other articulators. The 3D geometry of the various non-rigid organs is thus modeled as the weighted sum of a small number of linear components or DoFs. These components are 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 shape of the organs.

The corpus consisted of a set of artificially sustained articulations designed as to cover the maximal range of French articulations: the oral and nasal vowels [a ɛ e i y u o ø ɔ œ ɑ̃ ɛ̃ œ̃ ɔ̃], the consonants [p t k f s ʃ m n ʁ l] in three symmetrical contexts [a i u], and two specific articulations, a *rest* and a *prephonatory* position. Though limited to 46 French phonemes, this corpus proved to be sufficient for developing midsagittal articulatory models with nearly the same accuracy as corpora 40 times larger (Beautemps, Badin & Bailly (2001)).



Trans. Tech. Comm. Psychol. Acoust., The Acoustical Society of Japan, Vol. 36, No. 5, H-2006-77
Tech. Rep., SP2006-26, The Institute of Electronics, Information, and Communication Engineers, July 20, 2006

Note that our approach is *subject-oriented*, i.e. that the model is based on a single subject. This avoids merging the physiological characteristics and the coarticulatory or compensatory strategies that can vary fairly much between various subjects.

In the framework of our linear approach, one *DoF* may be defined for a given speech articulator as one variable that can control completely a specific variation of shape and position of this articulator, and that is *linearly* uncorrelated with the other DoFs over the set of tasks considered. We aim to exploit the correlations in the articulatory data to reduce the number of DoFs of the articulators. However, this approach must be carefully balanced by another criterion, the *biomechanical likelihood*, i.e. making sure that the DoFs are not related to control strategies actually used by the subject during the task but are really associated with movements that are plausible from the viewpoint of biomechanics.

Each linear DoF is *iteratively* determined by a mixture of PCA applied to part or whole of the organ region and of multiple regression of the data against control parameters either arbitrarily imposed such as jaw height or determined by the preceding PCA. Note that the solution of this type of linear decomposition is not unique in general: while PCA delivers optimal factors explaining the maximum of data variance with a minimum number of components, our approach allows some freedom to decide the nature and distribution of the variance explained by the components (for instance to make them more interpretable in terms of control), at the cost of a sub-optimal variance explanation and of weak correlation between components.

## 3. Determination of the organ shapes from MR and CT images

### *3.1. Preliminary remarks on speech organs anatomy*

The final result of an articulatory model is the shape of the complete vocal and nasal tracts, needed for simulating the aerodynamic / acoustic stage of the speech production process. The seemingly most straightforward way to deal with this problem was to develop *tract* or *duct* models (cf. e.g. Badin, Bailly, Raybaudi & Segebarth (1998)). However, this approach is not well suited for taking precisely into account the complex geometry of the various speech organ / cavities, e.g. tongue tip, sublingual cavity, velopharyngeal port or epiglottis. An *organ-based* modeling approach, where each organ and cavity are modeled separately, seems more appropriate, as it allows to reconstruct the oral and nasal tracts subsequently in more details, and to obtain more reliable area functions.

As the purpose of the study was not to develop biomechanical muscle models, but models of organs / cavities boundaries as a whole, attention was directed towards organ outlines behavior in regions where they define the vocal tract. Therefore, compromises were established: (1) different groups of muscle fibers, including sometime connective tissues, were grouped together; (2) boundaries that do not always correspond to actual organ boundaries were drawn in order to allow manipulating organs as closed volume objects. For example, the external part of tongue extrinsic muscles such as palatoglossus, styloglossus, or stylohyoid, muscles excluded; as well, only a part of the *levator veli palatini* muscle was included in the velum contours. Note that special care was devoted to distinguish tongue tip from mouth floor (see below). Similarly, the epiglottis was systematically excluded from the tongue contours, but will be modeled independently.

The following sections describe the methods used for obtaining the 3D surface representations of the various speech organs and cavities from MR and CT images.

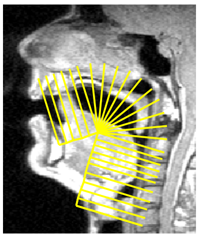


*Figure 1: Grid for transverse reslicing superimposed on a midsagittal image*

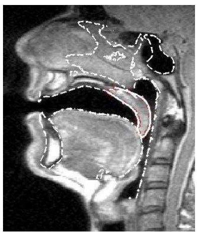


*Figure 2: Example of edited contours in the midsagittal plane*

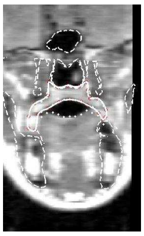


*Figure 3: Idem in a transverse plane*

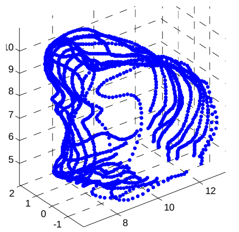


*Figure 4: Example of sagittal tongue contours for phoneme* $[l^a]$

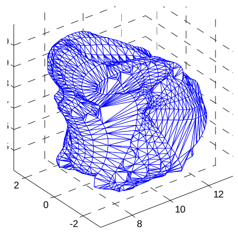


*Figure 5: Example of 3D tongue mesh for phoneme* $[l^a]$

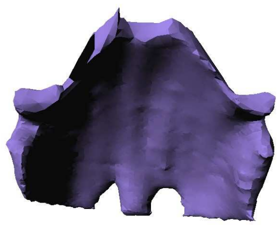


*Figure 6: Example of 3D velum surface*

### *3.2. Acquisition and pre-processing of the CT and MR images*

A Computer Tomography (CT) scan of the head of the subject was made, to serve as a reference. A stack of 149 axial images (512×512 pixels, 0.05 cm / pixel resolution, 0.13 cm inter slice space), was recorded for the subject at rest. These CT images, that provide a good contrast between bones, soft tissues and air, will be used to locate bony structures and to determine accurately their shapes for further use. Stacks of sagittal MR images were recorded for the subject sustaining artificially each of the 46 articulations of the corpus during about 35 sec. A set of 25 sagittal images (256×256 pixels, 0.1 cm / pixel resolution, 0.4 cm inter slice space) was obtained for each articulation. These images provide a good contrast between soft tissues and air, and also within soft tissues, but do not image clearly the bones.

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. The first step is thus to edit semi-manually the organ outline in each of the sagittal image where it can be seen. This is relatively straightforward for images close to the midsagittal plane, but much less so for images farther out from this midsagittal plane where the organ surface may be in tight contact with other structures or may just extend much further, such as the palatoglossus for instance. In order to improve the determination of the organ outline in these regions distant from the midsagittal plane, we created new images by reslicing the initial stack of images in planes having a more useful orientation, i.e. being more perpendicular to the organ surface. As a routine for the sagittal MRI stacks, we have used 27 transverse planes, i.e. the planes orthogonal to the sagittal ones, aligned on a semipolar grid (cf. Figure 1). The CT axial stack has been resliced into coronal and sagittal stacks.

### *3.3. Determination of the surface outline of the rigid bony structures*

A number of structures that makes up the vocal tract can be considered as rigid: jaw, hard palate, nasal passages, nostrils and various paranasal sinuses. The contours of these structures have been manually edited from CT images in planes with appropriate orientations (e.g. the nasal passages were edited from coronal and axial images). The set of all points forming the 2D planar contours was then expanded into the common reference 3D coordinate system. These 3D points were finally processed through a 3D meshing software (Geometrica Research Group at INRIA, http://cgal.inria.fr/Reconstruction) to form a 3D surface meshing based on triangles. Figure 8 that shows tongue nomogram for jaw height illustrates also the jaw surface.

### *3.4. Alignment of the images on a common reference*

Considering that the subject may have slightly changed position between two MR images stacks recording, it is important first to align all MRI stacks with a common reference. We use an arbitrary common 3D reference coordinate system attached to the skull of the subject. Each stack is aligned using an appropriate 3D transformation. This transformation, that corresponds to the six degrees of freedom of a solid object, 3 rotations and 3 translations, will be referred to as a (*3D*) *rototranslation*. It is obtained by aligning the rigid structures extracted from CT images (hard palate, nasal passages, paranasal sinuses) with the corresponding ones in the MR images stack, using a semi-automatic process: (1) anchor points of the rigid structures are manually marked with care on some of the MR images of the stack, (2) the appropriate 3D rototranslation is obtained by minimizing the cumulated distance between these points recast in 3D and the corresponding nearest points on the 3D rigid structures. A similar approach was proposed by Takemoto, Kitamura, Nishimoto & Honda (2004), the main differences being that their minimization error was the value of the volume overlap between the reference to align and the target data.

The previous procedure is also applied to the jaw for each articulation, to determine its relative position in relation to the fixed rigid structures: by combining this relative 3D rototranslation and the absolute one corresponding to the given stack, the jaw position is known in the common reference coordinate system.

### *3.5. Determination of the surface outline of the soft structures*

The determination of the soft structures (tongue, velum, nasopharyngeal wall) is achieved in much the same way as for the rigid structures, but from the MR images of each articulation. Planar contours were edited in both sagittal and transverse images. The contours resulting from the intersection of the rigid organs surface with the plane containing the image being edited were superposed on the image in order to provide bony anchor points not visible on MRI and thus very useful for the interpretation of the image. Figure 2 and Figure 3 illustrate the edition of the velum in both MRI stacks. In addition to bony structures, previously edited soft structures are superimposed on the image to help interpretation and maintain coherence. The contour (solid lines on Figure 2 and Figure 3) is established as a 2D b-spline curve controlled by a limited number of points. The set of all 2D plane contours expanded into the 3D coordinate system forms a 3D description of the given soft organ (cf. Figure 4).

### *3.6. Tongue tip and sublingual boundaries*

The sublingual cavity plays an important role in a number of articulations such as post-alveolar fricatives, laterals, back vowels and some consonants coarticulated with back vowels. As an illustration, note that the tongue tip is elevated enough to uncover the mouth floor – and thus to create a sublingual cavity – in 19 articulations of our corpus: [u o ɔ œ ã ɔ̃ m$^{u}$ ʃ$^{a}$ ʃ$^{u}$ p$^{a}$ p$^{u}$ k$^{a}$ k$^{u}$ l$^{a}$ l$^{i}$ l$^{u}$ ʁ$^{a}$ ʁ$^{u}$]. In these articulations, tongue tip and mouth floor contours can be easily be seen and traced (e.g. see Figure 1). For the other articulations for which the tongue tip rests on the mouth floor (see Figure 2), these contours are difficult to determine and have to be inferred indirectly.

In most of the central sagittal planes, three specific points can be identified on the tongue contours: (1) the extremity of the tip (referred to as *Tip*), (2) the rearmost point of the sublingual cavity, at the junction of the mouth floor outline and the inferior surface of the tongue tip (referred to as *Floor*), and (3) the point in the midsagittal plane where the mouth floor reaches the jaw internal surface, somewhere behind the lower incisors (referred to as *Jaw*).The *Tip* and *Jaw* points are easy to identify for all articulations. To simplify a little, we assume that all *Jaw* points in the sagittal planes are located in the same nearly axial plane orthogonal to the midsagittal plane, containing the midsagittal *Jaw* point and parallel to the plane tangent to the upper surface of the inferior teeth: the corresponding reference *jaw arch* is finally further simplified into a fourth order polynomial planar curve. The *Floor* point cannot be seen in articulations where the tongue tip rests on the mouth floor: it was thus decided to infer it for those cases, assuming its position being unchanging relatively to the jaw. We determined the reference *mouth floor arch* as the fourth order polynomial planar curve that approximates best the *floor* points of the [ʃ$^{a}$] articulation. The two reference arches defined as described above are supposed attached to the jaw, and can thus be positioned accordingly for each articulation, using the jaw rototranslation.

The strategy for editing the tongue contours in sagittal planes is finally the following: (1) position reference arches in 3D, find their intersection with the sagittal image plane and display these intersection points that are supposed to reflect the position of the *jaw* point and the *floor* point, (2) edit the contours as usual whenever visible, but interpolate using the *jaw* point and the *floor* point as anchor points when the tongue tip is touching the mouth floor. In addition, the three points (*Tip, Floor, Jaw*) were manually identified on each sagittal contour to allow an interpolation of the three sub-contours (mouth floor, tongue tip inferior surface, tongue superior surface) with a fixed number of points that will be used in the mesh matching process (see below). Similarly, the lower edge of the velum was determined and each sagittal contour interpolated with a fixed number of point for both upper and lower faces.

### *3.7. Fit of generic meshes to each articulation*

The plane contours defined above do not have a constant number of points and therefore cannot be used in a PCA type of analysis. In order to ensure a suitable geometric representation of all the articulations for a given soft organ, a unique generic 3D surface mesh, made of triangles, was defined for each organ and fitted by elastic deformation to each of the 46 3D shapes of the corpus. This procedure provides thus a 3D sampling of each organ surface with the same 3D vertices, which is appropriate for the development of the models. The elastic deformation of the generic mesh to fit each 3D shape is computed through a matching software developed by Szeliski & Lavallée (1996).

The lateral consonant [l[a]] was chosen as reference articulation for the tongue, as the tongue tip is sufficiently raised to let clearly appear the sub-lingual cavity, i.e. the lower surface of tongue tip and the mouth floor, in this articulation. Similarly, the nasal phoneme [ã] was chosen as a reference for the velum and nasopharyngeal wall, as the contact between velum and surrounding walls is minimum in this articulation. The meshing software was used to generate the generic meshes from the points sampling the various planar contours of the reference. As the reference and target shapes may be very different, it was judged more reliable to match the reference mesh progressively, using a series of ten intermediary sets of sagittal contours, where each point is linearly interpolated between its reference and target positions.

This matching process has finally resulted in a set of organ surfaces described in terms of triangular meshes having the same number of vertices, in a common reference coordinate system: 5239 vertices for the velum (RMS reconstruction error, 0.06 cm), 2110 vertices of the nasopharyngeal wall (RMS, 0.04 cm), and 1640 vertices for the tongue, for each of the 46 articulations of the corpus (RMS, 0.06 cm). This forms the basis for the articulatory modeling, as illustrated in the next section.

## 4. Articulatory models of velum, nasopharyngeal wall, jaw and tongue

### *4.1. Head tilt*

Head movements of the subject during the recording session could introduce a variation of the vocal tract shape as suggested by Kitamura, Takemoto, Honda, Shimada, Fujimoto, Syakudo, Masaki, Kuroda, Oku-uchi & Senda (2005). Head tilt changes in the sagittal orientation are the most important. They may increase the variability of the data without being related to a precise degree of freedom of the articulators and are therefore considered as a global articulatory degree of freedom of the head that is not specifically related to speech production. Therefore, it appears necessary, before modeling speech articulators independently, to remove the variability related to this degree of freedom from the data. Preliminary analysis has shown s strong correlation between spinal chord and nasopharyngeal wall. The horizontal coordinate of the back pharyngeal wall in the midsagittal plane at the height of uvula was used as tilt predictor, and referred to as *Head Tilt* parameter (HT).

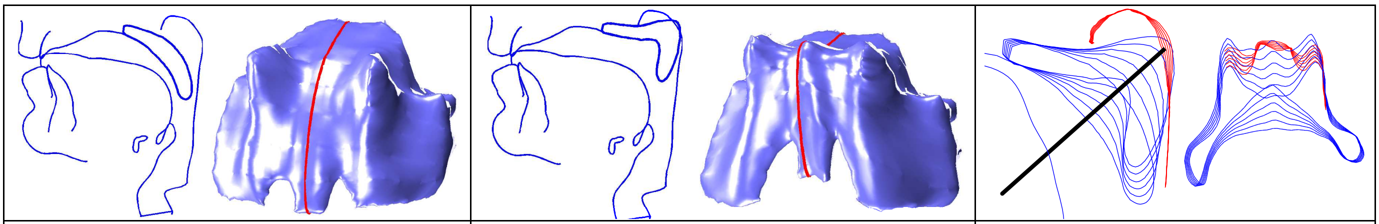
The contribution of this HT parameter to the explanation of the total variance of the data has been computed by means of linear regression: it explains 17% of the cumulated variance of the central part of the velum, 31% of the variance of the nasopharyngeal wall, and 7 % of the tongue surface. In the following, the models are built from the 46 velum and nasopharyngeal wall meshes cleared of the head tilt influence, i.e. once the linear contribution of the influence of HT parameter has been removed from the raw data. The influence of HT on the tongue was not deemed high enough to take it into account.

### *4.2. Articulatory model of velum of the velum and of the nasopharyngeal wall*

In order to determine the number and nature of the articulatory degrees of freedom of the velum, PCA was applied to the 2812×3 vertex coordinates that represent the surface of the central part of the velum over the 46 articulations corrected for tilt variations The first PCA parameter VL explains 83% of the cumulated variance of all the central velum points, while the associated RMS reconstruction error is 0.08 cm. This parameter can then be used to predict the entire velum surface based on a linear regression of the whole set of velum points on this parameter; in other words, the external regions of the velum can be predicted from the central region by means of this first PCA parameter. The effect of this parameter on the whole velum is illustrated on Figure 7 by the shape associated with the two extreme values of VL found in the data. The main movement associated with VL is a movement in an oblique direction. Considering its orientation and its prime importance for speech (Bell-Berti (1993)), the *levator veli palatini* can be thought to be much involved in this movement; this control parameter is thus referred to as *Velum Levator* (VL).

The second velum PCA parameter, VS, explains 6% of the total variance of the central points of the velum. The cumulated variance explained by the two parameters VL and VS attains 89% while the cumulated RMS reconstructions error lowers to 0.06 cm. This second parameter is much related to a horizontal displacement coupled with a vertical elongation of the velum, which complements the velopharyngeal port closure by a front to back movement and may significantly modify the velopharyngeal port constriction. Despite a relatively small gain in variance explanation, VS appears as a coherent complementary degree of freedom of the velum: (1) it reduces the RMS reconstruction error by 25%, (2) it may modify the velopharyngeal constriction in a meaning full way.

The velopharyngeal port closure is known to be controlled by the combined contribution of velum pulling and superior constrictor sphincter action (cf. e.g. Bell-Berti (1993)). Expectedly, we found that the VL parameter explains 47% of the total nasopharyngeal wall variance, with a reconstruction with an RMS error of 0.07 cm. The effect of VL in the midsagittal plane occurs in the upper region, along the main direction of velum deformation, but in an antagonist way. This movement corresponds to a variation of thickness of the Passavant's Pad, related to the contraction of the *palatopharyngeus* muscle and of the *pterygopharyngeal* portion of the superior constrictor, and participates to a global sphincter behavior of the velopharyngeal port. The parameter VS, that explains only 5% of the nasopharyngeal wall variance, was thus not retained in the nasopharyngeal wall model.



a) VL = -2 | b) VL = +1.4 | c) Nomograms of VL

*Figure 7: Velum shape for two extreme values of VL, and regular nomograms between these two values for the velum and the naso-pharyngeal wall in the midsagittal plane (c left) and in the transverse plane (c right) indicated by the transverse solid line on the left graph.*

In another study (Rossato, Badin & Bouaouni (2003)), one of the coils of an Electro Magnetic Articulograph was attached to the velum of the same subject uttering an extensive corpus of VCV sequences. We have determined the vertex of the velum mesh that was the closest to the velum coil position for all phoneme targets, and found a rather strong similarity between the domains occupied by the coil and the vertex in the midsagittal plane. This is an additional evidence of the validity of our data acquisition and modeling approach.

### *4.3. Jaw displacement model*

A strong correlation ($R^2$ = 0.93) was found between the lower incisor's height and the angle of rotation around a transverse axis, and smaller correlation with the incisor's horizontal position. The jaw height parameter *JawHei* has been used as a control parameter of the main jaw movement.

### *4.4. Tongue model*

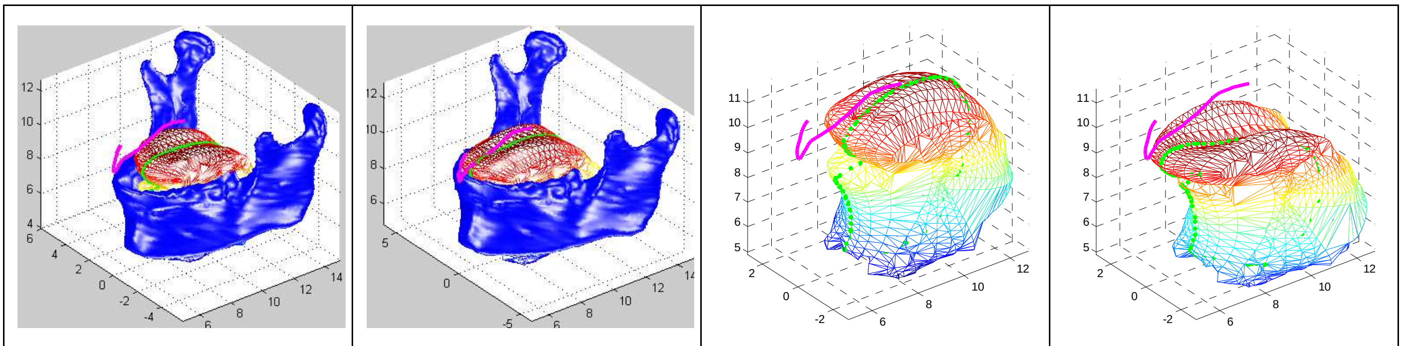
The first parameter, *jaw height JH*, is defined as the variable *JawHei* centered on its mean and normalized by its standard deviation. Its main effect is a tongue rotation around a point in the back of the tongue (see Figure 9). The next two parameters, *tongue body TB*, and *tongue dorsum TD*, are extracted by PCA from the coordinates of the midsagittal tongue contour, excluding the tongue tip region, from which the JH contribution has been removed. They control respectively the *front-back* and *flattening-arching* movements of the tongue (see Figure 9). The next two parameters, *tongue tip height TT*, and *tongue tip advance TA*, are extracted by PCA from the midsagittal tongue tip contour coordinates, from which the *TB* and *TD* contribution have been removed (see Figure 9).

Table I gives the variance, relative to the total variance of the full 3D coordinates, that is explained by each component, for both raw PCA and the controlled analysis described above. It appears that a amount of 82.6 % is explained by our controlled analysis, which is only 8% below the optimal result from a raw PCA with the same number of components.

Finally, the 3D tongue model is controlled by the five articulatory parameters *JH*, *TB*, *TD*, *TT1*, and *TT2*. The effects of these commands are demonstrated in Figure 8 which displays tongue shapes for two extreme values (–3 and +3) of one parameter, all other parameters being set to zero., and in Figure 9 which displays the midsagittal intersection of the 3D nomograms with sinusoidal variations of the control parameters.

| Raw ACP | | | Controlled ACP | | | |
|---|---|---|---|---|---|---|
| Parameter | varex | varexcum | Parameter | varex | varexcum | RMS(cm) |
| P1 | 0.5920 | 0.5920 | JH | 0.2221 | 0.2221 | 0.3452 |
| P2 | 0.1727 | 0.7647 | TB | 0.4139 | 0.6359 | 0.2361 |
| P3 | 0.0749 | 0.8396 | TD | 0.1172 | 0.7532 | 0.1944 |
| P4 | 0.0378 | 0.8774 | TT1 | 0.0303 | 0.7835 | 0.1821 |
| P5 | 0.0275 | **0.9049** | TT2 | 0.0425 | **0.8260** | 0.1633 |

*Table I: Evaluation of the tongue analysis: varex is the relative explained tongue data variance, varexcum its cumulated value, and RMS the RMS reconstruction error between the modeled mesh and the original planar contours.*



JH = +2 | JH = -2 | TB = -2 | TB = +2

*Figure 8 : Jaw and tongue positions for extreme values of JH and TD*

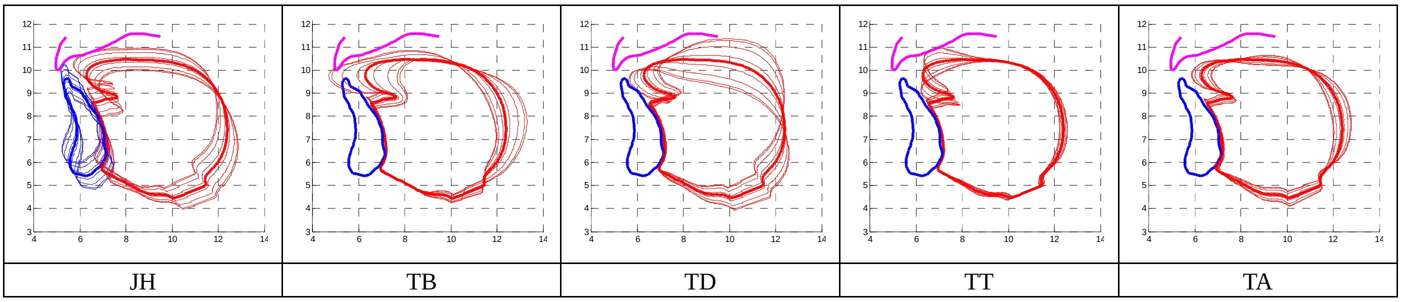



*Figure 9 : Projection in the midsagittal plane of the 3D nomograms*

## 5. Discussion and perspectives

The present work produced a number of valuable results. First, a database of 3D geometrical descriptions of tongue, velum and nasopharyngeal wall was established for a speaker sustaining a set of 46 French allophones covering the speech possibilities of the subject. Linear component analysis of these data revealed that five components could account for about 82 % of the total variance of the tongue shape, while 89 % of the velum variance is taken into account by two parameters, as well as 47 % of the nasopharyngeal wall (this is a smaller amount, but it should be noted that the absolute variance of the nasopharyngeal wall in about 10 times smaller than that of the velum itself).

Concerning the tongue, the five components correspond qualitatively to those found previously by Badin *et al.* (2002), on a more restricted corpus of 25 articulations only. The present study constitutes thus an extension of this study. It is also important to note that the variance explained here is about 10 % higher than in Badin *et al.* (2002), and that the tongue tip and sublingual cavities are much better defined. This will be useful for dealing with articulations such as post-alveolar fricatives, laterals, back vowels and some consonants coarticulated with back vowels. This is also one of the motivations for dealing with a 3D model. Another interest of this 3D model lies in the direct possibility to deal with tongue groove: Figure 8 illustrate well the strong groove variation related to the tongue dorsum parameter.

The velum / nasopharyngeal wall model constitutes an extension of the first model developed by Serrurier & Badin (2005, in press) on a restricted corpus of only 22 articulations.

The next modeling step will be the determination of the area functions for both oral and nasal tracts as a function of the shapes of tongue, velum and nasopharyngeal wall, and thus the determination of the acoustical characteristics of the complete tract. A longer-term objective is finally to extend all this process to the complete 3D vocal and nasal tracts in order to build a complete 3D articulatory model of speech.

## 6. Acknowledgements

We thank very sincerely Jean-François Lebas (Head of the Radiology Department, CHRU, Grenoble, France) for granting us access to the MRI equipment, Christophe Segebarth (Functional and Metabolic NeuroImagery Department, INSERM & UJF, Grenoble, France) and Monica Baciu (Psychology and NeuroCognition Department, UPMF, Grenoble, France) for managing the MRI recordings, Andreas Fabri (Geometrica team, INRIA, Sophia, France) for the letting us use his 3D meshing software, and Maxime Bérar (Image and Signal Laboratory, CNRS – INPG – UJF, Grenoble, France) for helping us with the elastic matching software.

## 7. References

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, pp. 249-254. Jenolan Caves, Australia.

Badin, P., Bailly, G., Revéret, L., Baciu, 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.

Beautemps, D., Badin, P. & Bailly, G. (2001) Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. Journal of the Acoustical Society of America, 109(5), 2165-2180.

Bell-Berti, F. (1993) Understanding velic motor control: studies of segmental context. In Phonetics and Phonolgy. Nasals, Nasalization, and the Velum (R.A. Krakow, editor), pp. 63-85. Academic Press Inc.

Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N. & Senda, M. (2005) Difference in vocal tract shape between upright and supine postures: observation by an open-type MRI scanner. Acoust. Sci. & Tech., 5, 465-468.

Rossato, S., Badin, P. & Bouaouni, F. (2003) Velar movements in French: an articulatory and acoustical analysis of coarticulation. In Proceedings of the 15th International Congress of Phonetic Sciences (M.-J. Solé, D. Recasens & J. Romero, editors), pp. 3141-3144. Barcelona, Spain.

Serrurier, A. & Badin, P. (2005, in press) Towards a 3D articulatory model of velum based on MRI and CT images. ZAS Papers in Linguistics.

Szeliski, R. & Lavallée, S. (1996) Matching 3-D anatomical surfaces with non-rigid deformations using octree-splines. International Journal of Computer Vision, 18(2), 171-186.

Takemoto, H., Kitamura, T., Nishimoto, H. & Honda, K. (2004) A method of teeth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. Acoustical Science and Technology, 25, 468-474.