A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data

Antoine Serrurier and Pierre Badin[a)]
GIPSA-lab, UMR 5216 CNRS-INPG-UJF-Université Stendhal, Département Parole and Cognition/ICP, 46 avenue Félix Viallet, 38031 Grenoble Cedex 01, France

(Received 1 June 2007; revised 21 January 2008; accepted 23 January 2008)

An original three-dimensional (3D) linear articulatory model of the velum and nasopharyngeal wall has been developed from magnetic resonance imaging (MRI) and computed tomography images of a French subject sustaining a set of 46 articulations, covering his articulatory repertoire. The velum and nasopharyngeal wall are represented by generic surface triangular meshes fitted to the 3D contours extracted from MRI for each articulation. Two *degrees of freedom* were uncovered by principal component analysis: first, VL accounts for 83% of the velum variance, corresponding to an oblique vertical movement seemingly related to the *levator veli palatini* muscle; second, VS explains another 6% of the velum variance, controlling a mostly horizontal movement possibly related to the sphincter action of the *superior pharyngeal constrictor*. The nasopharyngeal wall is also controlled by VL for 47% of its variance. Electromagnetic articulographic data recorded on the velum fitted these parameters exactly, and may serve to recover dynamic velum 3D shapes. The main oral and nasopharyngeal area functions controlled by the articulatory model, complemented by the area functions derived from the complex geometry of each nasal passage extracted from coronal MRIs, were fed to an acoustic model and gave promising results about the influence of velum movements on the spectral characteristics of nasals. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2875111]

PACS number(s): 43.70.Bk, 43.70.Jt, 43.70.Aj [CHS] Pages: 2335–2355

I. INTRODUCTION

According to Crystal (1997), "*nasality* is a term used in the phonetic classification of speech sounds on the basis of manner of articulation: it refers to sounds produced while the soft palate [or velum] is lowered to allow an audible escape of air through the nose." Understanding the production of nasal sounds therefore requires a good knowledge of the variable shape of the velopharyngeal port that connects the rigid nasal tract to the vocal tract, and that is delimited by the velum and the nasopharyngeal wall. A large number of studies have been devoted to the production of nasality (see, e.g., Ferguson, Hyman, and Ohala, 1975 or Huffman and Krakow, 1993).

A first gross estimation of the nasal tract geometry was proposed by House and Stevens (1956) from anatomical considerations. The first systematic anatomical measures of the nasal tract that we know of were performed by Bjuggren and Fant (1964), who traced cross-sectional contours from slices cut from a plastic mold of the nasal passages of a cadaver. The cross-sectional contours and nasal passage areas that they proposed have served as a standard reference for many decades, and have been used for acoustical simulations. The first, and as far as we know the only, sets of transversal images of the velopharyngeal port obtained by x-ray tomography were recorded by Björk (1961) for ten subjects uttering sustained articulations. Associating these images with sagittal x-ray tomography images of the same subjects, he found a linear relation between the nasal tract transverse coupling area and the velum/pharyngeal wall sagittal distance in the midsagittal plane for distances greater than 0.2 cm. The magnetic resonance imaging (MRI) technique is still considered to be the only imaging technique that is safe for the subject and that delivers comprehensive three-dimensional (3D) data. It has thus been largely used for determining the geometry of the vocal tract in speech (see, e.g., Baer, Gore, Gracco, and Nye, 1991, Story, Titze, and Hoffman, 1996; Engwall and Badin, 1999; Badin, Bailly, Revéret, Baciu, Segebarth, and Savariaux, 2002) and has allowed new measurements on live subjects and permitted researchers to obtain more accurate area functions of the nasal tract. In 1992, Matsumura and Sugiura (1992) published the first cross-sectional profiles of nasal passages from MRI images. Area functions derived from these measurements were thus proposed by Matsumura, Niikawa, Shimizu, Hashimoto, and Morita (1994) two years later. At the same time, Dang, Honda, and Suzuki (1994) led a similar study which proposed new area functions and compared their results with those obtained by Bjuggren and Fant (1964). They highlighted in particular the importance of mucosa in the nasal passages. Demolin, Lecuit, Metens, Nazarian, and Soquet (1998) subsequently performed a unique 3D study of velopharyngeal port opening from MR images recorded on four subjects pronouncing French nasal vowels and their oral counterparts; more details on the cross-sectional contours and areas of velopharyngeal port were then provided by Demolin, Delvaux, Metens, and Soquet (2003). Delvaux, Metens and Soquet (2002) studied the position and shape of the velum and the associated coordination of other articula-

[a)]Author to whom correspondence should be addressed. Electronic mail: pierre.badin@gipsa-lab.inpg.fr

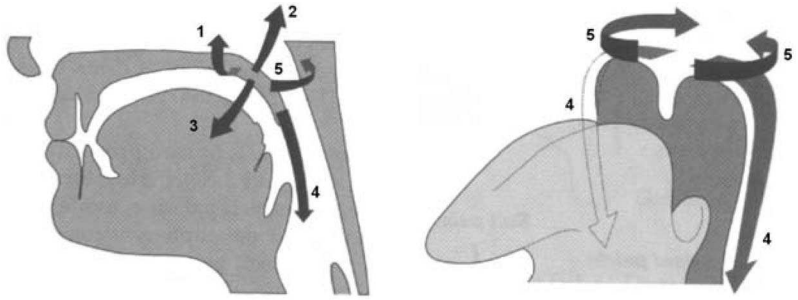


FIG. 1. Midsagittal view (a) and oblique anterior view (b) of vocal tract and schematic directions of action of the principal muscles involved in the velum and velopharyngeal port movements (from Kent, 1997); 1: Tensor veli palatini; 2: Levator veli palatini; 3: Palatoglossus; 4: Palatopharyngeus; 5: Pharyngeal superior constrictor.

tors such as tongue movements used by French speaking subjects in the production of French nasal vowels. Interestingly, they noted a possible contact between the velum and the tongue for low velar positions.

Physiologically, the velopharyngeal port is organized in a complex way. A network of muscles linking the surrounding organs, i.e., the velum, the lateral and posterior pharyngeal walls and the tongue, controls the velopharyngeal port's opening/closing mechanism. The velum, the principal organ involved in the mechanism, is known to be controlled mainly by five muscles (see Fig. 1). Its major muscle, the *levator veli palatini*, stretches symmetrically from the medial region of the velum to the right and left Eustachian tubes. The two other muscles of the velum are the *tensor veli palatini*, stretching laterally and symmetrically from the medial region of the velum to the base of the cranium and passing through a tendon acting as a pulley to ensure a lateral tensing of the velum, and the *uvulae muscle* (not visible in Fig. 1), located entirely in the uvula—an appendix of the velum in the midsagittal region (see, for example, Fig. 2 for various uvula positions), which is believed to have only a small impact on the velopharyngeal mechanism in speech. In addition, the

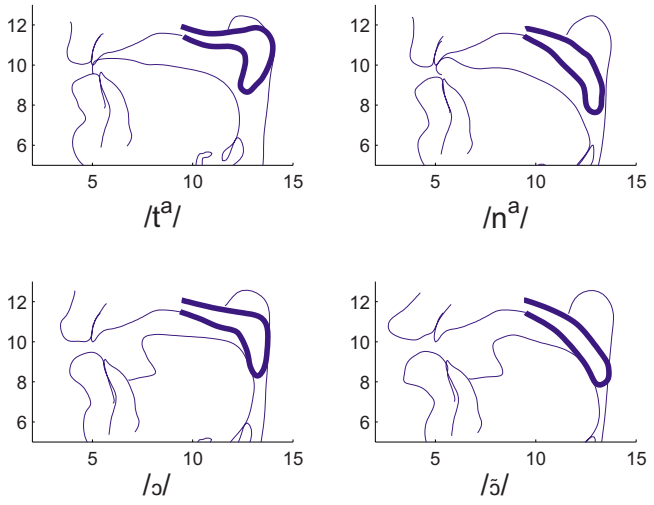


FIG. 2. (Color online) Midsagittal contours of the vocal tract for oral and nasal stop consonants /t^a/ and /n^a/ (top) and for oral and nasal vowels /ɔ/ and $\tilde{ɔ}$ (bottom). The thicker lines represent the velum contours.

velum is connected with its two neighboring organs: first with the tongue, through the *palatoglossus* muscle, with origin in the medial lower part of the velum and linking the lateral basis of the tongue along the borders of the oral cavity, known as the *anterior faucial pillars*; second with the pharyngeal walls, through the *palatopharyngeus* muscle, with its main origin in the medial upper part of the velum and linking the pharyngeal walls by forming the two *posterior faucial pillars* on both sides of the oral cavity (see, for example, Kent, 1997, for more detailed description of these muscles). The pharyngeal walls are principally active through the *superior, middle and inferior constrictor* muscles that surround the tract. The muscular structure of this region, in particular the interspersion between muscles from the velum, the pharyngeal walls, and the tongue, leads to a sphincter-like behavior (Amelot, Crevier-Buchman, and Maeda, 2003). Note that the contraction of the fibers of the *palatopharyngeus* muscle with those of the *pterygopharyngeal* portion of the superior constrictor leads to a prominence of the posterior wall called *Passavant's pad* (Zemlin, 1968), which contributes also to the sphincter effect. This effect may be speaker dependent, and at least four velopharyngeal closure patterns, depending on the anatomy of the speaker, have been reported (see, for example, Kent, 1997, and Amelot *et al.*, 2003, from fiberscopic data). The active or passive role played by each muscle involved in the closure mechanism during speech has led to various interpretations (see, for example Dickson and Dickson (1972); Bell-Berti (1976); Kollia, Gracco, and Harris (1995) and Wrench, 1999), although "the *levator (veli)palatini* muscle is widely accepted as the muscle primarily responsible for closing the velopharyngeal port by exerting an upward and backward pull on the velum" (Bell-Berti, 1993).

Due to the complex organization of the velopharyngeal port, the relative difficulty collecting geometric information in this region of the vocal tract, and consequently of measuring velopharyngeal movements, only a few articulatory models deal with nasals. House and Stevens (1956) proposed a basic model of nasal tract to oral tract coupling where the coupling seems to be implemented simply through a linear interpolation of the area function from the first velopharyngeal cross-sectional area to the first nasal tract area considered as fixed; they used this model for acoustical simulations

and perceptual studies of nasality. Fant (1960) investigated the influence of nasal area coupling in terms of acoustics by modeling the area function of the velopharyngeal port by a single tube. Maeda (1982) and Fant (1985) used a model similar to that of House and Stevens (1956), augmented with sinus cavities, in order to assess the contribution of these sinuses to the overall acoustic characteristics of nasals. Mermelstein (1973) proposed a crude geometric midsagittal model of velum shape and assumed the velar opening area to be proportional to the square of the distance between the current uvula position and the position attained when the velopharyngeal port is closed. This model has been used in particular by Teixeira, Vaz, Moutinho, and Coimbra (2001) for perceptual tests of synthesized Portuguese nasals (Teixeira, Moutinho, and Coimbra, 2003).

The development of more realistic models of speech production—and particularly of nasals—calls for more detailed 3D articulatory models of the velopharyngeal port and of the nasal cavities. Indeed, the accurate area functions of the complex nasal passages and velopharyngeal port that are needed to feed acoustical models, and thus to generate speech, cannot be obtained with simple models: for some nasal articulations, e.g., the French back nasal vowels, as highlighted by Demolin *et al.* (2003), the uvula can be in contact with both the back of the tongue and the pharyngeal wall in the midsagittal region (see Fig. 2), leading to a midsagittal occlusion, though the channels on each side of this occlusion remain open. Such articulations thus require a 3D description. More or less successful ad hoc transformations from midsagittal shape to area function have been proposed for the oral tract (see, e.g., Sundberg, Johansson, Wilbr, and Ytterbergh, 1987; Beautemps, Badin, and Bailly, 2001); but the only model proposed for the velopharyngeal port (Mermelstein, 1973) cannot deal with a midsagittal occlusion. It thus appears that a 3D model in which appropriate information is provided about the transverse structure of the vocal and nasal tracts is clearly needed.

This present study is intended to result in a nasal tract that complements the 3D linear articulatory models previously built in our laboratory (Beautemps *et al.*, 2001; Badin *et al.*, 2002) in the framework of the development of talking heads (Badin, Bailly, Elisei, and Odisio, 2003). Specifically, we attempted to reconstruct 3D nasal cavities, velum and nasopharyngeal wall shapes from MRI images from one subject uttering a corpus of sustained French articulations, and to develop a corresponding 3D linear articulatory model. This *organ-based* approach, as opposed to the *tract* approach that cannot take into account the complex geometry of the various speech articulators, aims in particular to explore the articulatory degrees of freedom of the articulators, following the approach of Badin *et al.* (2002) to modeling of the tongue and lips, based on the same French subject and the same corpus.

The following sections present the various articulatory data acquired on the subject, their analysis in terms of uncorrelated linear articulatory degrees of freedom, and the associated linear articulatory models. A preliminary acoustical evaluation of this articulatory model is also presented. This study constitutes an extension of the 3D articulatory modeling of nasals initiated in Serrurier and Badin (2005a) and Serrurier and Badin (2005b).

II. ARTICULATORY DATA

A. Subject and speech material

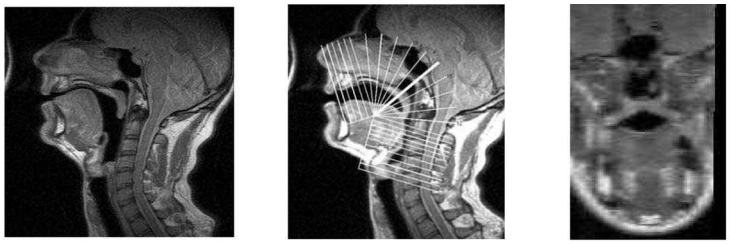
Designing a corpus and recording appropriate data obviously constitutes the first important stage of a data-based approach to articulatory modeling. As the principle underlying linear modeling is that any articulation should be decomposable into a weighted sum of basic shapes, that constitutes a minimal basis for the space of articulations, the corpus should constitute a representative sampling for this space. One way to achieve this is to include in the corpus all articulations that the subject can produce in his language. The corpus thus consisted of: the 10 French oral vowels [a ɛ e i y u o ø ɔ œ], the four French nasal vowels [ɑ̃ ɛ̃ œ̃ ɔ̃], the artificially sustained consonants [p t k f s ʃ m n ʁ l] produced in three symmetric contexts [a i u], and, finally, a "rest" position and a "prephonatory" position. These last two are produced without sound, lips open, nasal tract connected to the oral tract, jaw open, in a neutral position for the rest articulation and in a position ready to phonate for the prephonatory articulation. Altogether, there are 46 target articulations. This limited corpus proved to be sufficient for developing midsagittal articulatory models with nearly the same accuracy as corpora 40 times larger (Beautemps *et al.*, 2001). This corpus will be referred to as the *main corpus*.

As the present study constitutes the first attempt to elaborate a 3D articulatory model from MRI data, only one subject was considered: we chose the male French speaker already involved in the development of a midsagittal articulatory model based on a cineradio-film (Beautemps *et al.*, 2001), and of 3D models of tongue, lips, and face based on MRI and video data (Badin *et al.*, 2002). He was about 1.65 m tall and 43 years old at the time of recording the main corpus.

B. Data

As highlighted in the introduction, one of the most efficient and accessible methods of collecting 3D sets of vocal tract shapes, and considered to be safe for the subjects, is magnetic resonance imaging. Following Badin *et al.* (2002), the present study is based on 3D sets of MR images collected for each articulation of the main corpus, i.e., 46 stacks of sagittal images, from which 3D shapes of the soft organs are extracted. However, due to the difficulty of distinguishing air from bones in MRI, a set of computed tomography (CT) scans of the subject at rest was also recorded to serve as a reference and to help interpret the MR images.

Other data were also collected on the same subject for specific purposes. The geometry of the nasal passages being very complex and air passages sometimes very narrow, a set of coronal images considered to be perpendicular to the direction of the nasal tract has been recorded in order to optimize air/tissue detection. In order to complement the MRI



a) Sagittal – Original image – [i] b) Re-slicing grid – [a] c) Reconstructed transverse image – [a]

FIG. 3. Examples of MR images for [i] and [a] articulations (a and b) and (c) of a transverse image for [a] reconstructed along the thick white line in (b).

static shapes with dynamic data, electromagnetic midsagittal articulatory (EMA) data have also been recorded.

1. Sagittal MRI

Stacks of sagittal MR images were recorded using the 1 Tesla MRI scanner Philips GyroScan *T*10-NT available at the Grenoble University Hospital. The subject was instructed to sustain the articulation throughout the whole acquisition time, approximately 35 s for each of the 46 articulations. 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 25×25 cm, a thickness of 0.36 cm, and an inter-slice center to center distance of 0.4 cm was obtained for each articulation. The image resolution is 0.0977 cm/pixel, approximated to 0.1 cm/pixel in the rest of the article, the images size being 256×256 pixels. From these images it was possible to make the distinction between soft tissues and air, and to discriminate the soft tissues, but not to clearly distinguish the bones. They have thus been used to collect the 3D shapes of soft organs, but the CT scans were required to identify the bony structures. Note that the subject was in a supine position, which may alter somehow the natural shape of articulators (cf. Tiede, Masaki, and Vatikiotis-Bateson, 2000, and Kitamura, Takemoto, Honda, Shimada, Fujimoto, Syakudo, Masaki, Kuroda, Oku-uchi, and Senda, 2005). Examples of midsagittal images for the vowels [i] and [a] are shown in Figs. 3(a) and 3(b).

2. CT images

A stack of 149 axial images with a size of 512×512 pixels, a resolution of 0.05 cm/pixel, and an inter-slice space of 0.13 cm, spanning from the neck to the top of the head, was recorded by means of a Philips Mx8000 scanner for the subject at rest (see one example image in Fig. 4(a)). These images allow us to distinguish bones, soft tissues and air, but do not allow for the identification of different soft tissues. They have been used to locate bony structures and to determine accurately their shapes for reference (see Sec. II C 2).

3. Coronal MRI

A stack of 32 coronal images with a size of 256×256 pixels, a resolution of 0.1 cm/pixel, and an inter-slice space of 0.4 cm, spanning from the atlas bone to the tip of the nose, was recorded for the subject at rest (see Fig. 5(a)) to optimize detection of the nasal cavities.

4. EMA

Dynamic data were collected through an electromagnetic midsagittal articulograph (Rossato, Badin, and Bouaouni, 2003). One of the coils of the articulograph was attached to the velum about halfway between the hard palate-velum junction and the tip of the uvula, so as to provide a robust estimation of velum movements (see Fig. 13(a)). The corpus consisted of all the combinations of nonsense words [VCV], V being one of the 14 French oral or nasal vowels and C one of the 16 French consonants [b d g p t k v z ʒ f s ʃ m n ʁ l].

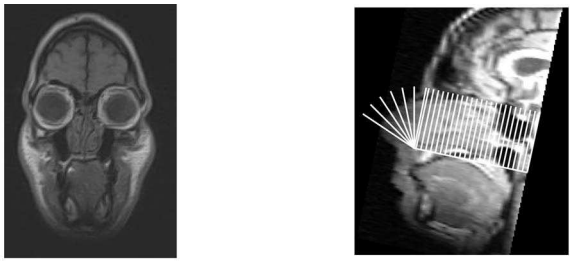
C. Preprocessing of images

1. Re-slicing of the original image stacks

Due to the complexity of the contours of the various organs, the relatively low resolution of the images, and the need for an accurate reconstruction of the organs, extraction of the contours was performed manually, plane by plane. This is a rather reliable process, except for regions where the surface of the structure is tangent to the plane, and thus difficult to trace and not very accurate. For instance, while the

a) Axial – original image b) Reconstructed coronal image c) Reconstructed sagittal image

FIG. 4. Example of original (a) and reconstructed (b and c) CT images.



a) Example of original coronal MR image b) Semipolar grid of nasal tract sampling

FIG. 5. Original coronal MRI located between the atlas bone and the beginning of the nose (a), and semipolar grid showing the location of the original and reconstructed images (b).

pharyngeal wall is easy to trace in the midsagittal plane, the sidewalls of the pharynx are nearly tangent to the off-midline sagittal planes. This is why we have supplemented the original stacks of images used for the articulatory model with a single orientation (*axial* for CT images and *sagittal* for MR images) by extra sets of images reconstructed by intersection of the initial stack with planes having a more useful orientation, i.e., being more perpendicular to the organ surface.

The CT images have thus been resliced/interpolated in two stacks of 512 coronal images and 512 sagittal images, leading altogether to three stacks of perpendicular CT images with a high resolution of 0.05 cm/pixel in the three orientations (see Fig. 4).

For the sagittal MR images, the initial sagittal stack was resliced in images perpendicular to the vocal tract, considering that they will be used to extract organ shapes delimiting the vocal tract (e.g., velum, tongue, etc.). They were thus re-sliced in 27 planes orthogonal to the midsagittal plane and intersecting it along a semipolar grid, as illustrated in Fig. 3(b). Each new image was given a resolution of 0.1 cm/pixel. Finally, we disposed of two redundant orthogonal stacks of MR images for each articulation.

In the same way, sagittal MR images have been re-sliced in the nostril region along a polar grid, as shown on Fig. 5, in order to follow the nearly 90° bend of the nasal passages near the nostrils, and thus to maintain the images perpendicular to the nasal tract and therefore optimally adapted to border tract detection. Eight new images of arbitrary size 256×256 pixels, which have the same 0.1 cm/pixel resolution as the original images, have thus been created.

The original coronal images, being already perpendicular to the tract between the cavum and the beginning of the nostrils, were used as is.

2. Extraction of bony structures and alignment of the image stacks on a common reference

Regarding the data processing, a contour drawn on a specific image (such as the solid line on Fig. 9(a) which represents the velum) can be expressed in three different types of units: (1) two-dimensional (2D) coordinates in the image plane of the contour in a continuous pixel unit (i.e., subpixel resolution), referred to as *2D pixels*; (2) 2D coordinates in the image plane in cm, which can be transformed into continuous pixel units (and vice versa) by using the image pixel resolution parameter, referred to as *2D cm*; and (3) 3D coordinates in cm, which can be transformed into *2D cm* and vice versa using the knowledge of the image plane 3D location, referred to as *3D cm*. The following process refers to these various coordinates depending on the requirements.

As the subject's head may have moved between the recordings of the various articulations, it was necessary to align the image stacks of each articulation on a common reference framework, before attempting to determine the shape of the soft structures.

Fixed bony structures such as the cranium, hard palate, and various paranasal sinuses (maxillary, sphenoidal and frontal sinuses) that can be seen—at least partially—on each image, were used as landmarks. The set of 3D triangular meshes of the surface outlines of these structures was made from the stack of CT images where bones appear clearly, and were then used as a common reference.

This process was implemented in the following way. Manual segmentation of each organ outline was done plane by plane, in one of the three CT stacks, or a combination of them, depending on the shape and orientation of the organ, in such a way as to maximize the accuracy of complex organs (e.g., the maxillary sinuses were manually segmented in the coronal stack while the sphenoid sinus was segmented in the sagittal one). The planar contours are edited as 2D splines controlled by a limited number of points; for example, Figs. 9(a) and 9(b) illustrate (with a solid line) the contour of the velum manually segmented from MR images of articulation $[l^a]$. The whole set of 2D planar contours, whose coordinates are originally in 2D pixels, were then expressed in 2D cm and expanded into 3D cm to form a set of 3D planar contours representing the specific organ (see Fig. 9(c) for the set of 3D planar contours of the velum for the configuration $[l^a]$). These 3D points have been then processed through a 3D meshing software (Fabri, 2007) to form a 3D surface mesh based on triangles (see further Fig. 10 for the velum and pharyngeal wall).

The alignment of each stack of images on the common framework defined by the fixed bony structures is performed in three steps. First, the rigid fixed bony structures (hard palate, nasal passages, paranasal sinuses, etc.) are chosen as an arbitrary common reference. An absolute 3D reference coordinate system, attached to the skull of the subject, is arbitrarily defined as follows: (1) the x axis is oriented from anterior to posterior in the midsagittal plane and approxi-

mately in the occlusal plane, the y axis from left to right, and the z axis from feet to head; (2) the lower edge of the upper incisors in the midsagittal plane is considered as the origin of the coordinate system. Second, the stacks of CT images are manually aligned with this reference as a rigid body; thus, the associated transformation gives the position of the three stacks of CT images in the common reference coordinate system. This transformation corresponds to the six degrees of freedom of a rigid body and is thus defined by six parameters: three parameters for the 3D rotation and three parameters for the 3D translation; it will be referred to as a (3D) *rototranslation*. Finally, as the subject may have slightly changed position between two MRI stacks recordings, each stack was aligned with the common reference by using the appropriate 3D rototranslation. This rototranslation was determined by aligning specific rigid structures (hard palate and paranasal sinuses), extracted from CT images, with each stack of the MR images. This alignment is a semiautomatic process: (1) anchor points of the rigid structures are manually marked with care on some of the MR images of the stack, (2) the associated 3D rototranslation is determined by a simple minimization of the added distance between these 3D points (originally 2D planar points expressed in 2D pixels, then expressed in 2D cm and finally expanded in 3D cm) and the corresponding nearest points on the 3D rigid structures. A similar approach was proposed by Takemoto, Kitamura, Nishimoto, and Honda (2004), the main difference being that their minimization error was the value of the volume overlap between the reference to align and the target data.

The same alignment procedure was also applied to the jaw for each articulation, in order to determine its relative position in relation to the fixed rigid structures: by combining these relative 3D rototranslations and the absolute one corresponding to the given stack, the positions of these two structures are known in the common reference to each articulation.

Note that this procedure provides the geometric transformation between the 3D reference coordinate system, in which contours and meshes are expressed in 3D cm, and the stacks of images, in which contours are expressed in 2D pixels. This transformation can obviously be considered in both direct and reverse directions, the two sets of coordinates being thus equivalent. Note, however, that in order to maximize the accuracy of the contours detection, we decided to draw contours (in 2D pixels) on original images whenever they existed, and to use this geometric transformation to produce planar contours in 3D cm in the reference coordinate system, rather than applying the transformation to an image and drawing contours in 3D cm on this transformed image that is necessarily more noisy.

D. Nasal passages: 3D geometry and area function

1. Contours extraction

The nasal passage contours were manually segmented on the 24 coronal MR images and the eight resliced images defined by the grid visible in Fig. 5(b). The corresponding set of contours for the nasal tract is illustrated in Fig. 6. Three different regions can be considered in the nasal tract: (1) the cavum, which is the cavity located above the velum and behind the septum wall of the nasal passages (its cross-sectional contours are visible on the slices numbered from −1.2 to 0 cm in Fig. 6), (2) the choanae, which form the middle region of the nasal tract from the cavum to the nostrils (their cross-sectional contours are visible on the slices numbered from 0.4 to 6.4 cm in Fig. 6), and (3) the nostrils, which are the outlet of the nasal tract (their cross-sectional contours are visible on the slices numbered from 6.8 to 9.2 cm in Fig. 6).

The complex shape of the nasal passages, particularly in the middle region, combined with the relatively low resolution of the images (0.1 cm/pixel) makes it difficult to trace the nasal passage contours. In order to improve the accuracy of detection, the contours of the nasal passages were extracted from the CT images in the same way along the same grid, and were superimposed on the MR images. This provided a useful help for the determination of nasal passages, as soft tissues/air boundaries are better defined in CT images. We noticed, however, that the areas extracted from the CT images were about twice as large as those extracted from the MR images (cf. Fig. 7). This may be ascribed to the fact that nasal mucosa have low tissue densities and thus may not be visible on CT images. Figure 7 illustrates the resulting cross-sectional areas extracted from CT and MR images.

These results are in overall agreement with the observations made by Dang *et al.* (1994) and complement their study of four subjects. Note that the volume of the cavum is highly dependent on the position of the velum, which is very low in our case (see Fig. 5(b)); a low velum position results in large cross-sectional areas, as can be seen in Fig. 7 for slices located on the back of the septum, i.e., having negative abscissa. Note also the asymmetry between the left and right passages and the irregularities between two consecutive sections. These characteristics are related to the rather intricate shape of the nasal passages which are made up of many thin and interlaced partitions (see Fig. 6 that illustrates the nasal passage contours). Each nasal passage constricts mildly at a point located between the choanae and the nostrils. This constriction of about 1 cm^2 is located around 7 cm from the separation point of the nasal tract into choanae and 2.4 cm from the nostrils' outlet (at $x=6.6$ cm). The size of this constriction has often been discussed in the literature: it varies from 0.23 cm^2 for House and Stevens (1956) to about 2 cm^2 for Bjuggren and Fant (1964); Feng and Castelli (1996) estimated it to range between 0.5 and 1 cm^2 and fixed it at 0.6 cm^2 for their acoustic simulations; based on measurements from MR images, Dang *et al.* (1994) showed constrictions of about 1 cm^2, and Matsumura *et al.* (1994) between 0.5 and 1 cm^2. Ultimately, the significant difference between areas extracted from CT and MR images emphasizes the importance of mucosa in the nasal passages.

2. Area function determination

Oral and nasal tracts can be approximated in speech by a succession along a horizontal line of cylindrical tubes whose lengths and cross-sectional areas represent more or less finely the real tract; the area function provides this cross-sectional area as a function of the abscissa along the midline

FIG. 6. (Color online) Cross sectional contours of the nasal tract along the grid lines shown on the bottom right subplot: cavum (from −1.2 to 0 cm), left and right choanae (from 0.4 to 6.4 cm) and left and right nostrils (from 6.8 to 9.2 cm).

of the tract. An area function for the nasal passages has been computed from the cross-sectional contours extracted along the grid, assuming that the cavities are geometrically sampled in each image plane. The elementary tubes making up the area function are then assumed to be located between two adjacent sections and are determined as follows.

Each section area is calculated from the contours in each image plane: areas corresponding respectively to multiple *outer* contours (i.e., for which tissues are outside and air inside) are accumulated (see, e.g., the two contours defining the left nasal passage on slice 3.6 cm in Fig. 6), and conversely areas corresponding to *inner* contours (i.e., for which tissues are inside and air outside) are subtracted (see, e.g., the two contours defining the interior limitations of the tract on slice 0 cm in Fig. 6). The center of the tract in this section is calculated as the average of the center of gravity of each contour, weighted by the corresponding area, counted as positive for outer contours and negative for inner contours.

A multiplicative correction factor is applied to the area in order to compensate the bias introduced by the fact that the vocal tract is usually not intersected by a plane that is exactly orthogonal to its midline. This oblique section actually increases the section area by a factor which is the cosine of the angle between the ideal plane that would be orthogo-

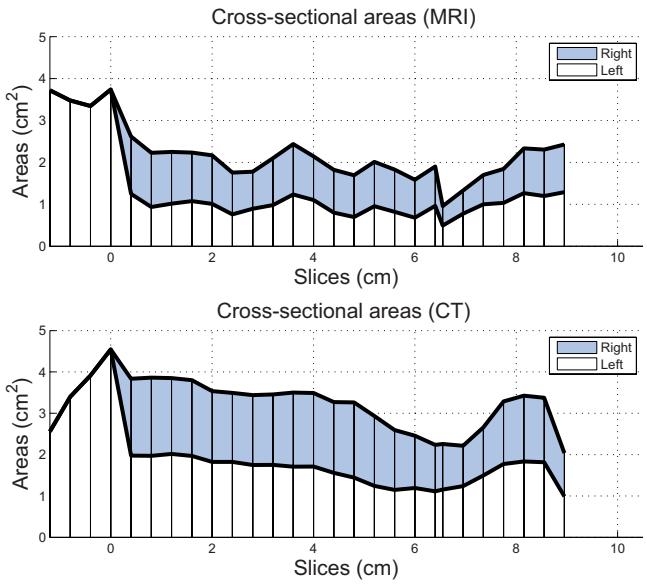


FIG. 7. (Color online) Cross-sectional areas of the nasal tract extracted from MR (top) and CT (bottom) images: cavum, left and right passages areas. The x axis represents the distance in cm from the cavum to the nostrils. The origin is arbitrarily chosen as the first image plane where nasal tract splits into two choanae separated in the midsagittal plane by the septum wall; the gap between each point is fixed to 0.4 cm, which is exactly the distance between two consecutive images in the back and middle part, and an extrapolation of this distance to the front part.

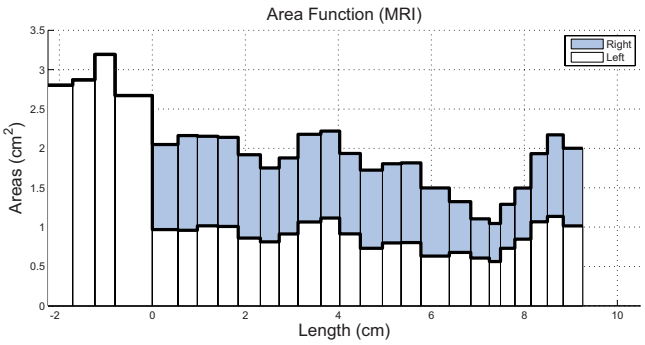


FIG. 8. (Color online) Area function of the nasal tract: each tube is represented by a length on the x axis and an area on the y axis.

nal to the midline and the actual cutting plane. The correction factor is thus computed as the cosine of the angle between (1) the 3D vector normal to the *actual* cutting plane of the tract and (2) the 3D vector normal to the *ideal* cutting plane of the tract. The tract midline is approximated locally by a circle passing through the center of gravity of the cross-sectional contour actually computed and those of its two neighboring sections, and thus the 3D vector normal to the *ideal* cutting plane is the tangent to this circle in the section considered.

The area of the tube is finally computed as the mean of the corrected areas of the two adjacent sections.

The length of the tube is computed as the distance between the 3D coordinates of the centers of the tract in each of the two sections.

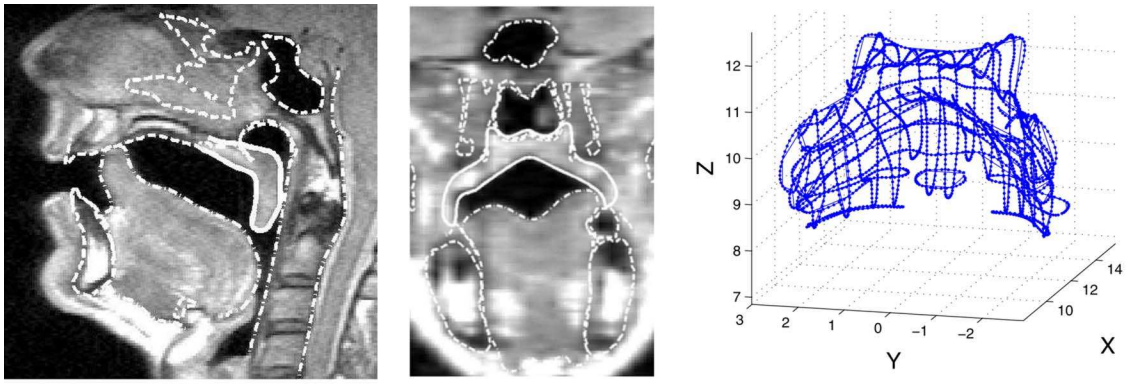
3. Nasal passage area functions

The area functions of the nasal passages derived from the MR contours are illustrated in Fig. 8. Each elementary tube is represented by a line whose length is plotted on the x axis and whose area is plotted on the y axis. We observe a length of about 9.2 cm for the nasal passages, from the beginning of the septum (at $x=0$ cm) to the final outlet from the nostrils, and a constriction of about 0.6 cm long and 1 cm^2 large at 1.7 cm from the outlet of the nostrils (at $x=6.9-7.5$ cm).

In comparison, the length of the nasal passages proposed by House and Stevens (1956) is around 8.5 cm long and the area decreases from 5.9 to 0.23 cm^2 at the nostrils, which is considered as the constriction. If the cross-sectional contours proposed by Bjuggren and Fant (1964) seem similar to our data extracted from the MR images and the length of 7–8 cm for nasal passages appears slightly smaller, their areas seem largely overestimated, from more than 7 cm^2 in the middle part to around 2 cm^2 for the nostrils, which constitutes the constriction. Dang *et al.* (1994) highlighted the importance of the mucosa in the nasal passages by comparing the nasal contours of the same subject in a normal condition and in a condition where the nasal mucosa had been spread with a vaso-constrictor substance, and concluded that Bjuggren and Fant overestimated the areas due to a retraction of mucosa on the cadaver. This assumption is confirmed in this study, where cross-sectional areas extracted from CT images appear overestimated, likely due to the low density mucosa not visible on these images. The areas reported by Matsumura *et al.* (1994) appear significantly lower than those of Bjuggren and Fant (1964) (less than 3 cm^2 for the middle and front part) and the constriction appears between 0.8 and 2.8 cm from the tip of the nose, depending on the subject. Finally, as discussed above, our results correspond globally with those reported by Dang *et al.* (1994).

E. Determination of the soft structures

The shapes of the soft structures, i.e., the velum and the nasopharyngeal wall, were determined for each articulation in much the same way as the rigid structures, using both original sagittal MR image stacks and associated sets of reconstructed images. In addition, the 3D rigid structures meshes, properly aligned with the MR image stacks as described above, were then intersected by the planes corresponding to the MR images, leading to 3D planar contours in centimeters, easily expressed in 2D pixel planar contours, and superimposed on the MR images in order to provide some useful anchor points for the interpretation of the images and for the tracing of the soft structure contours (see dashed lines in Figs. 9(a) and 9(b)). Figure 9 illustrates the segmentation of the velum in both MRI stacks for the consonant [l^a]. In addition to the bony structures, previously manually segmented soft structures other than the velum are also superimposed on the images to help the detection of the velum, and to maintain coherence in tracing the various organs (see the tongue contours previously segmented in dash-dotted lines in Figs. 9(a) and 9(b)). The definition of the outline of a soft organ is of prime importance for the 3D shape extraction process. Air/tissue boundaries are obvious borders and are generally easy to detect on MRI (see, e.g., Fig. 9(a) where the velum contour appears clearly). As our primary goal is modeling the vocal tract boundaries, the borders of organs that are never in contact with air are not very important but may, however, play an important role in the elastic deformation of the organ under consideration (as discussed later in this section). Contour tracing in these regions has thus been performed according to the principle of coherence: the border choice aims to ensure coherence across configurations, as well as anatomic likelihood. This difficulty arises particularly for the superior part of the velum in lateral regions on both sides of the velopharyngeal tract (see Fig. 9(b)). Indeed, the velum shape was designed as a closed surface (except at the connection with the hard palate) and consequently the planar contours must be closed as well. The back and lateral pharyngeal wall outlines of the velopharyngeal port are on the contrary designed as a single open surface from the uvula level to the beginning of the cavum. Contours have therefore been traced in such a way as to stick on the velum contours in lateral regions in order to ensure closure of the tract. The set of all 2D planar contours traced in this way, initially expressed in 2D pixels, then in 2D cm, expanded into 3D cm and finally aligned with the common reference by using the 3D rototranslation of the corresponding images stack (see Sec. II C 2), forms a 3D description of the given soft organ (see, e.g., Fig. 9(c)).



a) Segmentation in a sagittal plane

b) Segmentation in a transverse plane

c) Set of all 2D planar contours

FIG. 9. (Color online) Example of manual segmentation (a and b), and of 3D representation (c) of the velum for a [l^a] articulation. In a and b, the dashed lines correspond to bony structures superimposed on the images while dash-dotted lines correspond to soft structures other than the velum edited. The solid line corresponds to the manually edited velar contour.

As linear analysis methods such as principal component analysis or multiple linear regressions require each observation to bear on the same number of variables, it was necessary to ensure a common geometric representation of each soft organ for all the articulations in the corpus. A unique generic 3D surface mesh, made of triangles, was thus defined for each organ. These generic meshes were then fitted by elastic deformation to each of the 3D shapes of the corpus, providing a 3D representation of each organ surface with the same 3D vertices. The generic meshes were derived from the set of 3D planar contours of the [$\tilde{a}$] articulation by means of a 3D meshing software (Fabri, 2007). This articulation presents the advantage of a minimal contact between the velum and the surrounding structures, and thus is the most efficient way to extract the full 3D surfaces. Furthermore, the generic meshes (shown in Fig. 10) were slightly smoothed with the help of the *SmoothMesh* software (Huber, 2007) in order to reduce the noise introduced by the meshing procedure and the discrepancies due to the general problem of sampling 3D surfaces by a relatively small number of planar contours. The fit to each 3D shape extracted was computed by means of the matching software *TestRigid* developed at the TIMC laboratory in Grenoble (Couteau, Payan and Lavallée, 2000), and based on the elastic deformation of the generic meshes to the target configuration (see example of such a target in Fig. 9(c)). In order to ensure convergence of the matching algorithm, the deformation from generic source meshes to target configurations was supervised. About ten intermediate targets were obtained by linear interpolation from the planar contours which were achieved by intersections of the source mesh with planes of the target contours to the target contours themselves. The matching procedure was then applied in stages by introducing these targets one after the other. This method avoids unrealistic deformations that may occur when fitting the source mesh to too disjoint a target. The deformations of the soft organs obtained in such a way are believed to be close to the real deformations, i.e., the vertices of the mesh can be considered as flesh points and their relative positions coherent across configurations. This reconstruction process finally provides a set of soft organ surfaces described in terms of triangular meshes having the same number of vertices for all the articulations of the corpus and expressed in a common reference coordinate system. For all these articulations, each of the 46 closed velum surfaces is com-

a) Anterior view of the velum

b) ¾ posterior view of the velum

c) ¾ posterior sagittal view of the left-half pharyngeal wall

FIG. 10. (Color online) Three-dimensional generic meshes of the velum (a and b) and of the back pharyngeal wall (c) reconstructed for the [$\tilde{a}$] articulation.

posed of 5239 vertices and each of the 46 pharyngeal wall surfaces from the uvula level to the beginning of the cavum is composed of 2110 vertices. This matching process has resulted in a root mean square (RMS) reconstruction error between the fitted mesh and the target points over the 46 configurations of 0.06 cm for the velum and of 0.04 cm for the pharyngeal wall.

These sets of 3D data form the basis of the articulatory modeling of the subject speech organs, as will be described in the next sections.

III. 3D LINEAR ARTICULATORY MODELS OF VELOPHARYNGEAL PORT

A. Principles of articulatory modeling

As emphasized by Kelso, Saltzman, and Tuller (1986), the speech organs are made of a large number of neuromuscular components that offer a potentially large dimensionality and which must be functionally coupled in order to produce relatively simple gestures. Thus, following the approach used by Beautemps *et al.* (2001) and Badin *et al.* (2002), we consider that these simple gestures for a given organ can be represented in terms of *independent degrees of freedom* defined as the specific displacements and deformations of the organ which are linearly uncorrelated with the other degrees of freedom of this organ over the considered set of tasks. These degrees of freedom can be determined by observing the correlations between the various parameters that constitute the accurate geometrical description of the articulators' shapes and positions, and retaining only independent parameters. These correlations stem from a series of implicit or explicit constraints: physical continuity of the articulators, biomechanical constraints, and the nature of the task in relation to the control.

The movements of a speech organ can be modeled at two levels: a *physical* level, which deals with the displacements and deformations of the organ as *biomechanical* responses to muscle actions, and a *functional* level, which deals with the geometrical representation of the displacements and deformations of the organ. Linear articulatory modeling aims to determine the functional degrees of freedom of the organs and their associated control parameters. The correlations between the various movements observed on the articulator over the set of considered tasks are exploited to reduce the number of degrees of freedom and to determine the associated components. However, this approach is carefully used with the biomechanical likelihood constraint: the components must correspond to plausible movements in terms of biomechanics and must not result from a pure control strategy of the subject.

The linearity of the model means that the shape of an articulator for a given articulation is expressed as a linear combination of the components of the model, weighted by the set of control parameters corresponding to the articulation. In other words, assuming that the shape of an articulator is defined by the set of the coordinates of all its points means that varying a single input value corresponding to a specific component, while keeping all others constant, leads to a displacement of each point of the articulator along a straight line.

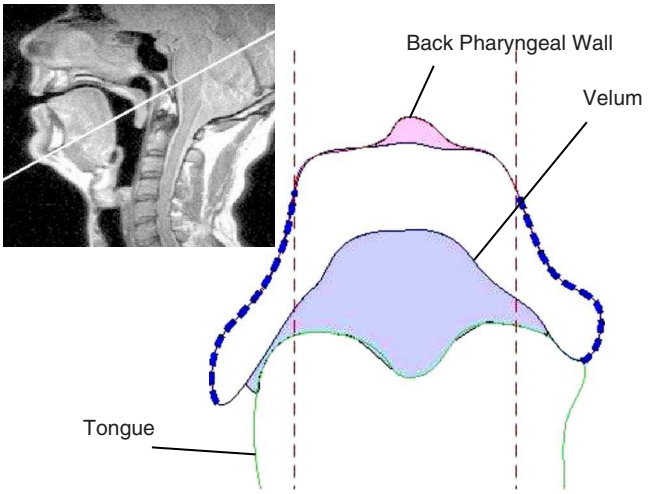


FIG. 11. (Color online) Cross-sectional contours of the velum, the pharyngeal wall and the tongue in the transverse plane shown on the MRI in the upper left corner for an [i] articulation. Oral and nasal cross-sectional areas are marked by a slightly dark color. The vertical dashed lines show the lateral limits of the analysis region of the velum at 1.5 cm on both sides of the midsagittal plane; the dashed contours of the velum beyond these vertical lines are considered to be irrelevant for the study.

Before describing the articulatory model based on this approach of the velopharyngeal port, i.e., of the velum and nasopharyngeal walls, we define the selection of the data used for the analysis and assess the contribution of head tilt to the movement of these articulators.

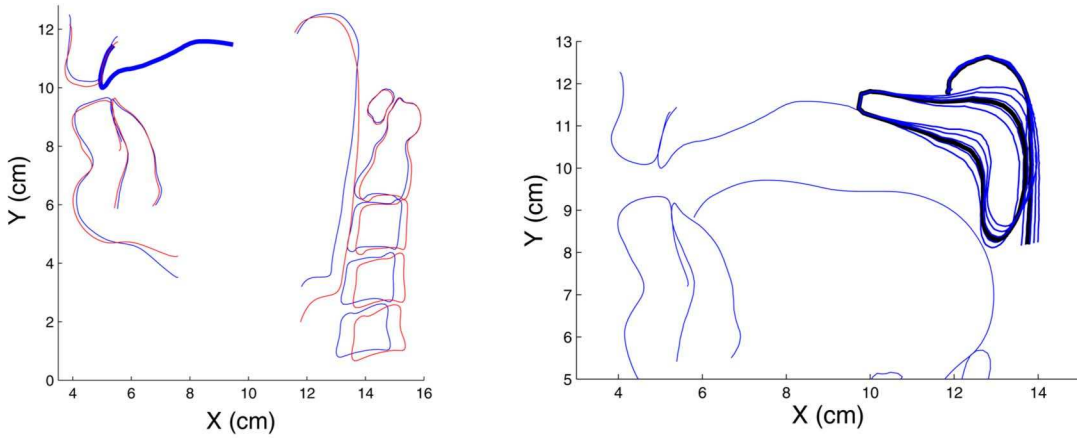
B. Selection of data for analysis

As explained above, the planar contours of the organs have also been traced in regions that do not bring information about the 3D shape of the vocal tract. The main objective of this study is to model the boundaries of soft organs that contribute to the 3D vocal tract shape: the organs' contours in these regions were thus considered irrelevant for the model and not included in the analysis.

According to combined criteria of anatomic likelihood and coherence across articulations, some regions of the velum considered as irrelevant for the model have, however, been traced, as illustrated in Fig. 11 by dashed lines. In order to simplify the choice of relevant regions, we applied a conservative approach and restricted the analysis to a region no further than 1.5 cm on both sides from the midsagittal plane (cf. Fig. 11). As will be explained in Sec. III E, the model was based on the analysis of this central part of the velum composed of 2812 vertices. In addition, the external regions will be shown to be very well correlated with the central region, which is in the end not surprising, considering the fact that the muscular fibers of the velum stretch continuously from the center to the more external regions.

C. Head tilt influence

Considering that the subject's head may have moved during the recording session, the stacks of images have been



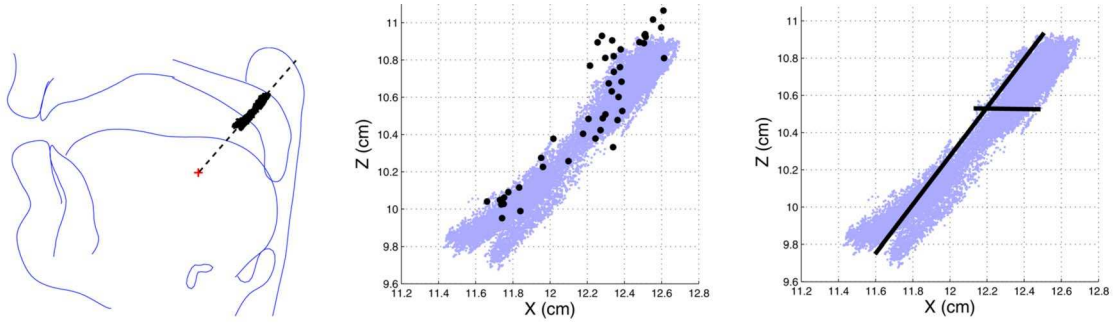
a) Superposition of contours in a sagittal image for [i] and [t^a] articulations

b) Nomogram of HT parameter for [a] articulation

FIG. 12. (Color online) Influence of the head tilt on organ shapes. Superimposition of contours in the sagittal image plane closest to the midsagittal one for the articulations of the 3D corpus corresponding to extreme tilts [i] and [t^a] (a). (b) Trace in the midsagittal plane for an [a] articulation of the 3D nomogram of the velum and pharyngeal wall for the HT parameter between the two extreme tilts of the data, −1.5 and 2.5.

aligned with a common reference system attached to the cranium (see Sec. II C 2). However, a change of head tilt, that produces rotation of the head within a sagittal plane, as for example when the chin and nose rotate down closer to the chest, introduces a variation of the vocal tract shape suggested by Kitamura *et al.* (2005) which is not taken into consideration in this alignment. Such a head movement may increase the variability of the data without being related to a precise degree of freedom of the articulators and is therefore considered to be 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: the global component associated with the tilt is computed on the whole vocal tract and its contribution removed from each organ's data. Anatomically, a tilt change that affects the pitch of the head corresponds to a rotation of the cranium structure in a sagittal plane around the cervical vertebrae apex to which it is attached through various ligaments. As the cranium structure is aligned with a fixed coordinate system in the current study, a change of head tilt creates a rotation of the spine in the reference system and an associated deformation of the vocal tract in the pharyngeal region and of the midsagittal trace of the posterior longitudinal ligament. The most posterior line in Fig. 9(a) illustrates the midsagittal trace of posterior longitudinal ligament, attached to cervical vertebrae. The deformation of this latter trace—also obviously related to the cervical vertebrae position—can thus be considered as a good marker of the head tilt variation, though the posterior pharyngeal wall would be a more convenient marker. Therefore, the relationships between vocal tract posterior pharyngeal wall and posterior longitudinal ligament have been investigated using a set of midsagittal MRI recorded for three typical positions of the head (normal, forward and backward tilt), for the same subject on a subset of 38 phonemes of the main corpus. The recording conditions were similar to those of the sagittal MRI recordings presented in Sec. II B 1. The velum, the posterior longitudinal ligament and the back pharyngeal wall were manually segmented, from glottis to cavum, on the images of this set, and aligned with the cranium in the midsagittal plane. As could be expected, it was found that the first Principal Component Analysis (PCA) component of the pharyngeal wall coordinates is strongly correlated with the first PCA component extracted from the posterior longitudinal ligament, for a corpus including large tilt movements (correlation of 0.99), as well as for a corpus where there are only small movements, corresponding to our data (correlation of 0.83). The first PCA pharyngeal wall component was thus considered to be a good head tilt predictor, as illustrated on Fig. 12(a). Finally, we verified that the horizontal coordinate of the back pharyngeal wall in the midsagittal plane at the height of the uvula, the lower limit of our analysis data, was correlated with this first pharyngeal wall component; it was used as a tilt predictor, and referred to as *Head Tilt* (HT) parameter.

The contribution of this parameter to the explanation of the total variance of the data has been computed by means of linear regression of the central part of the velum and of the pharyngeal wall on this parameter. It was found that 17% of the accumulated variance of the 2812×3 variables of the velum and 31% of the 2110×3 variables accumulated variance of the pharyngeal wall were explained by the HT parameter. The effect of this parameter in the midsagittal plane is demonstrated on Fig. 12(b). The nasopharyngeal wall movement appears to be coherent with a pure head tilt movement: it moves along in a horizontal direction, maximally in the lower region and remains fixed in the upper region. The velum movement corresponds to a simultaneous motion along horizontal and vertical axis which suggests a correlation with the principal movement measured through the EMA coil (cf. Fig. 13(a)). Indeed, since the shape of the vocal and nasal tracts changes under an independent action of the head tilt, the subject may introduce a compensatory action to correct for the tilt influence and thus attain the phonetic target. The fairly constant midsagittal distance be-



a) Midsagittal space covered by the EMA coil for the EMA corpus

b) Superimposition of MRI data positions on EMA data

c) Superimposition of nomograms of VL (oblique line) and VS (horizontal line) on EMA data

FIG. 13. (Color online) Midsagittal space covered by the velum EMA coil for the corpus of VCV sequences (a); superimposition on these EMA data of the *X-Z* coordinates of the *pseudo-EMA vertex* of the velum mesh: for the 46 configurations of the main corpus (b), for a movement associated with VL (oblique line on c) and to VS (horizontal line on c).

tween the velum and the back pharyngeal wall suggests an active compensation by the subject, logically related to the principal movement of the velum.

The following models are built from 46 velum and pharyngeal wall meshes for which the linear contribution of the HT parameter has been removed from the raw data.

D. Articulatory degrees of freedom of a fleshpoint on the velum

Before exploring the various articulatory degrees of freedom of the velum and back pharyngeal wall 3D surfaces, it is worth analyzing the articulatory degrees of freedom of the EMA velum coil presented in Sec. II B 4. Figure 13 displays the whole midsagittal space covered by the coil for the corpus of VCV sequences produced by the subject. This figure clearly reveals two main displacement directions: one dominant direction along an oblique axis over a range of about 1.6 cm, and a secondary direction along the horizontal axis over a range of about 0.35 cm. This observation suggests two articulatory degrees of freedom for a fleshpoint in this region of the velum located halfway between hard palate and uvula.

E. 3D linear articulatory model 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, principal component analysis (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 accumulated 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 with a very slight increase of global RMS reconstruction error of less than 0.01 cm. The effect of VL on the whole velum is illustrated on Figs. 14(a) and 14(b) by the shape associated with the two extreme values of VL found in the data and on two planes by the re-sliced contours for a regular succession of VL values between these two extremes (Fig. 14(c)). The main movement associated with VL is a movement in an oblique direction similar to that observed in the EMA data. Considering its orientation and its prime importance for speech (Bell-Berti, 1993), the *levator veli palatini* muscle can be thought to be much involved in this movement; this control parameter is thus referred to as *Velum Levator*. Moreover, the deformation of the velum in the midsagittal plane (Fig. 14(c) left) as it is raised from a low flat position to a high horizontal position broken at a right angle in its middle suggests a mechanism attached to the middle of the velum and pulling at about 45° degrees, the rest of the velum following this movement passively. This deformation could be ascribed to the *levator veli palatini* contraction.

The second velum PCA parameter, VS, explains 6% of the total variance of the central points of the velum. The accumulated variance explained by the two parameters VL and VS attains 89% while the accumulated RMS reconstruction error lowers to 0.06 cm. This second parameter, whose effects on the whole velum are displayed on Fig. 15, is 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 (see Sec. IV B).

The third PCA parameter explains less than 3% of the full variance and corresponds to a slight movement in the left-right direction that has no coherent explanation and can be considered to be an inaccurate parameter. It was thus not retained for the model.

It has been shown in the literature (Amelot *et al.*, 2003) that the velopharyngeal port closure is controlled by a combination of velum pulling and superior constrictor sphincter action. A strong correlation between pharyngeal wall and velum could thus be expected for our data. It was first shown by linear regression applied to the 2110×3 variables of the

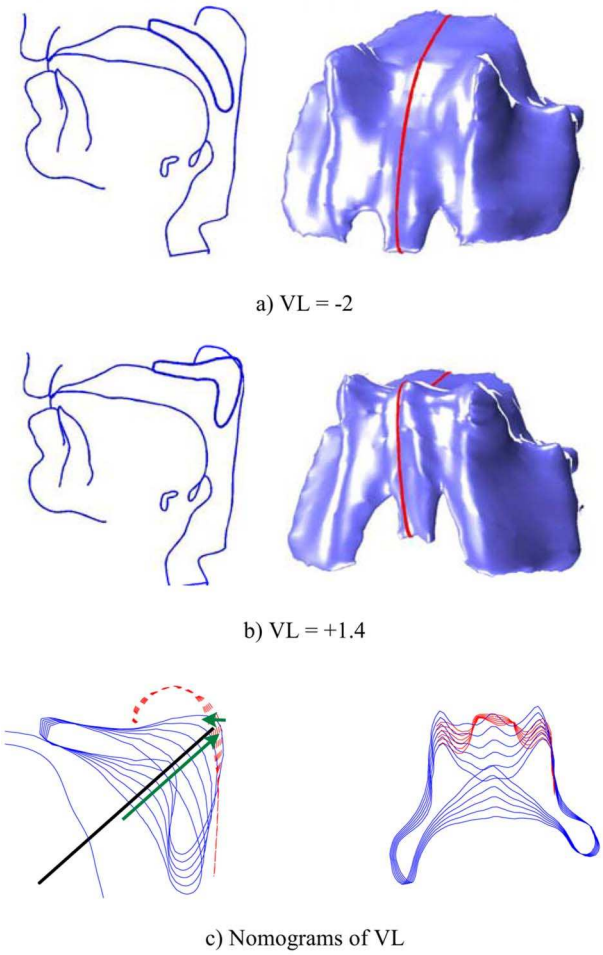


FIG. 14. (Color online) Posterior views of velum model 3D meshes (a and b right) and corresponding midsagittal contours (a and b left) for two extreme values of parameter VL (−2 in a and +1.4 in b); regular nomograms between these two values for the velum and the pharyngeal wall (c) displayed in the midsagittal plane (c left) and in the transverse plane (c right) indicated by the transverse solid line on the left graph (in the same way as in Fig. 11).

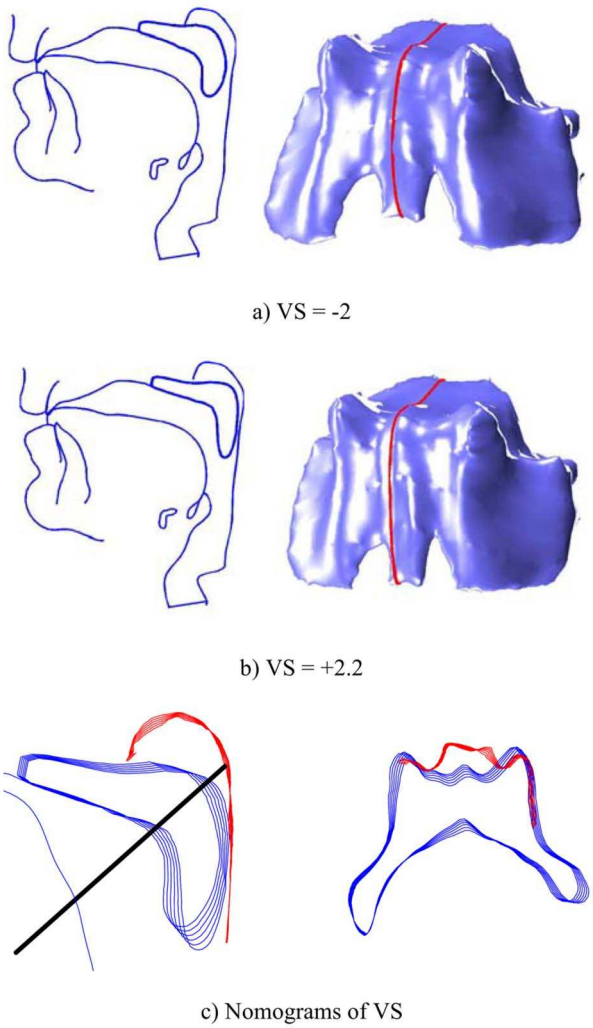


FIG. 15. (Color online) Posterior views of velum model 3D meshes (a and b right) and corresponding midsagittal contours (a and b left) for two extreme values of parameter VS (−2 in a and +2.2 in b); regular nomograms between these two values for the velum and the pharyngeal wall (c) displayed in the midsagittal plane (c left) and in the transverse plane (c right) indicated by the transverse solid line on the left graph (in the same way as in Fig. 11).

pharyngeal wall over the main corpus, that the VL parameter explains 47% of the total pharyngeal wall variance and allows a reconstruction with an RMS error of 0.07 cm. The validity of the contribution of this parameter to the pharyngeal wall variance explanation was further confirmed by a PCA applied to the pharyngeal wall: the first PCA parameter, which explains 50% of the total pharyngeal wall variance, is correlated with VL with a coefficient of 0.96. This confirms the strong relation between velum and pharyngeal wall and justifies a combined velum/pharyngeal wall model—i.e., a velopharyngeal port model—controlled by the VL parameter. The effect of VL on the pharyngeal wall can be seen in Fig. 14(c). The associated deformation in the midsagittal plane occurs in the upper region, along the main direction of velum deformation, but in an antagonist way, i.e., in the backward direction when the velum lowers and in the forward direction when the velum pulls up, so as to complement the velopharyngeal port closure. 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 pharyngeal constrictor*, and participates in a global sphincter behavior of the velopharyngeal port.

The parameter VS explains only 5% of the full variance and the associated deformation of the pharyngeal wall does not seem to correspond to a meaningful movement, as visible in Fig. 15(c). This parameter was thus not retained in the pharyngeal wall model. Note that the global variance of pharyngeal wall is about four times smaller than that of the velum surface; this means that 5% of the variance of the pharyngeal wall corresponds—in terms of movement—to much less than 6% of the velum surface variance.

Finally, note that for a number of articulations, the velum, and especially the uvula, is in contact with the tongue dorsum: the velum seems thus to undergo a mechanical deformation, pushed backward by the tongue more than if no contact was established (see, for example, Fig. 2 for [ɔ̃]). This is due to the fact that the tongue dorsum may be stronger than the uvula. In order to assess the possible effects of this contact, a model similar to the previous one was devel-

oped from the 28 of the 46 configurations of the main corpus that were free of velum/tongue contact. To complement this contactless model, the region of the velum in contact with the tongue was geometrically shifted in order to simulate the tongue push. This type of modeling did not bring significant improvement in terms of variance explanation and RMS error reconstruction and was thus not considered any further.

F. Consistency between MRI and EMA measurements

In order to establish more precisely the relation between the 3D velum model and the EMA velum coil coordinates, which represent the displacement of a real flesh point of the velum, the specific vertex of the velum mesh that could be considered as the flesh point on which the coil is attached was determined, according to the following method. (1) Each of the 46 articulations of the MRI corpus were linked with one or more occurrences of the same articulation in the EMA corpus; (2) for each occurrence, the velum vertices were sorted according to their distance from the corresponding EMA velum coil location; (3) the general rank of a given vertex was then defined as the maximum of its distance rank in each occurrence, that corresponds to the maximal distance over all the occurrences; (4) the *pseudo-EMA vertex* was finally chosen as the vertex with the smallest global rank, which ensures that it has the smallest of the maximal distances over the occurrences.

Figure 13(b) displays, on top of the midsagittal space covered by the EMA velum coil, the *pseudo-EMA vertex* position for the 46 shapes of the velum mesh constituting analysis data. The accurate correspondence between *pseudo-EMA vertex* and the real EMA velum coil confirms the consistency between the two sets of data.

Due to the linear nature of the velum model, each control parameter drives the displacement of each vertex along straight lines. Figure 13(c) displays such lines for the *pseudo-EMA vertex* as a function of VL and of VS in the midsagittal plane on top of the midsagittal space covered by the EMA velum coil. This confirms the excellent correspondence between the movements of the real EMA coil and of the *pseudo-EMA vertex* reconstructed from the MRI images, and thus the validity of our approach.

Finally, note that despite a relatively small gain in the amount of variance that is explained, VS appears as a coherent complementary degree of freedom to the velum: (1) it reduces the RMS reconstruction error by 25%, (2) it may modify the velopharyngeal constriction in a meaningful way, and (3) it helps the *pseudo-EMA vertex* trajectories to cover the 2D space of the real EMA data in a coherent way.

G. Recovery of 3D velum shape from midsagittal measurements

In the general context of speech production research, the relation between a 3D shape of an organ and its midsagittal contours constitutes an interesting issue. Badin *et al.* (2002) have shown that the 3D shapes could be predicted for the most part from midsagittal contours for the tongue, the face and the lips. Thus, it may be interesting to infer the time trajectories of the 3D velum and pharyngeal wall from midsagittal measurements such as EMA velum coil recordings. Therefore we assessed the possibility of recovering the 3D geometry of the velopharyngeal port from the position of the velum EMA coil in the *X-Z* plane by means of the model. We attempted to estimate the values of the two parameters VL and VS from the position in the *X-Z* plane of the *pseudo-EMA vertex* on the velum mesh for each of the articulations of the main corpus. As the coordinates of this point in the *X-Z* plane are linearly related to VL and VS by the velum model, the determination of these two parameters is straightforward. Note that the reconstruction error for this point is null, but may be non-optimal for the other vertices. A global RMS reconstruction error of 0.08 cm was finally observed over the full corpus of the 46 configurations for the model driven by VL estimated from the *pseudo-EMA vertex*; the error was down to 0.07 cm for the model driven by combined VL and VS. These results seem to be acceptable in comparison with the model RMS reconstruction errors of 0.08 and 0.06 cm for VL and VL plus VS, respectively. Moreover, the real VL values and their estimation from the *pseudo-EMA vertex* are correlated with a coefficient of 0.98; similarly, a correlation of 0.95 was found for VS. These results confirm the possibility of predicting the global 3D shape of the velopharyngeal port from midsagittal data and from a single point such as an EMA coil record in particular.

IV. ARTICULATORY MODELING AND AREA FUNCTIONS

The new articulatory model represents important knowledge about the articulatory behavior of the subject's velum and nasopharyngeal wall. However, since an aim of speech production modeling is to predict the speech signal, it is very important to be able to derive area functions from the 3D shape of the oral and nasal tracts in order to be able to characterize the resulting acoustics.

A. Oral and velopharyngeal area function determination

At this level, a complete articulatory model of the oral tract of the subject is not yet available. Thus, in order to assess the acoustic influence of the velum movements, the complete area functions of the oral and nasal tracts for various values of parameters VL and VS were needed. We have thus manually traced oral tract cross-sectional contours on all the transverse images defined by the semipolar grid illustrated in Fig. 3(b) for the three point vowels [a i u]. Note that the superimposition on the images of the bony structure contours, including the teeth as illustrated in Fig. 9(a), allows more accurate and reliable vocal tract contours to be drawn. For a given articulation and given values of VL and VS, the velum and pharyngeal wall have then been intersected with transverse planes in the velum region (see top left of Fig. 16), with an angle of 8° between two adjacent planes. As visible in Fig. 11, the nasal tract in such planes is delimited by the velum and nasopharyngeal wall contours, while the oral tract is delimited by the velum and tongue contours. Thus, the tongue contours have been manually traced in the

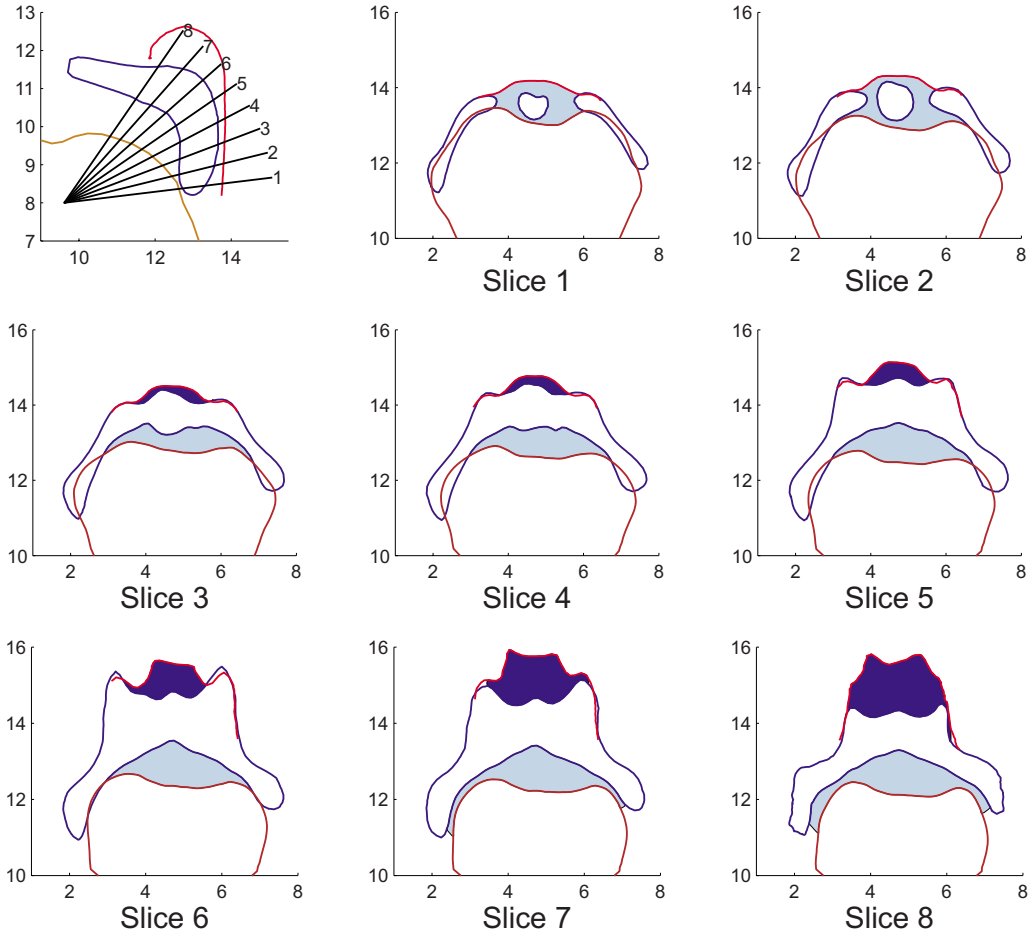


FIG. 16. (Color online) Display of cross sections of oral and nasal tracts for VL=0.3, VS=0, and for the tongue shape of [a] along the eight planes in the velum region shown by the midsagittal grid on the top left of the figure. Oral cross-sectional areas are displayed in clear and nasal tract areas in dark. The velum intersections in grids 1 and 2 correspond to isolated flesh areas inside the oral tract.

corresponding intersecting planes for [a, i, u]. Then, the oral and nasal tracts' cross-sectional contours in these planes have been automatically detected. This detection assumes necessarily that the contours are closed. Therefore, organs separated by less than 0.03 cm were considered to be in contact, which ensured that the associated tract contours were closed. This was particularly needed in regions where the tongue and velum were in contact. Note that in planes near the horizontal ones, the intersection with the velum—in fact the uvula—can be limited to a closed contour isolated inside the oral tract (see slices 1 and 2 of Fig. 16). In this case, the cross-sectional area of the inner contour must be subtracted from that of the outer contour, to take into account the reduction of tract area due to the presence of the uvula. Figure 16 displays the cross-sectional organ outlines and corresponding oral and nasal cross-sectional areas in the velum region for VL and VS equal respectively to 0.3 and 0 and for the tongue shape of [a].

Oral and nasal area functions have then been computed from the cross-sectional contours following the process described in Sec. II D 2. The first cross section with a distinct nasal tract (e.g., grid 3 in Fig. 16) is considered as the beginning of the nasal tract, and thus connected to the nasal passages described in Sec. II D 1 through the cavum. The eight oral tract cross sections corresponding to the velum region are integrated into the whole set of the vowel cross sections already traced, in order to obtain complete oral area functions from the glottis to the lips. Note that the outlet tube at the lips has been lengthened by 1.1 cm to take into account the labial horn, i.e., the region of lips ahead of the lip corner where the cross-sectional vocal tract outline is not a closed contour any longer (Badin, Motoki, Miki, Ritterhaus, and Lallouache, 1994). The next two sections describe the influence of the velum position on oral and nasal area functions.

B. Influence of velum movement on the velopharyngeal area function

Figure 17 displays nomograms of the velopharyngeal area function for variations of VL and VS corresponding to those of Figs. 14 and 15. As expected, variations of VL (Fig. 17(a)) induce important area function variations: the volume of the velopharyngeal tract from beginning of its separation from oral tract to the point of separation into two nasal passages varies from 9.70 cm^3 for a velum maximally open to 4.25 cm^3 at occlusion. The velopharyngeal port constriction is located close to its lower end. When the velopharyngeal port closes, the velum pulls up, while the uvula, the lower end of the velopharyngeal tract, and the constriction location

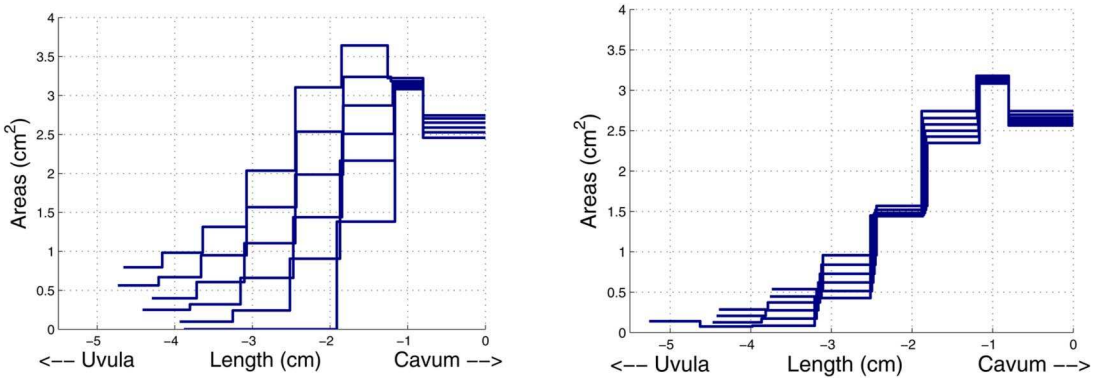


FIG. 17. (Color online) Variations of velopharyngeal port area function as a function of VL (a) and VS (b) parameters. The origin of the abscissa along the tract is arbitrarily chosen as the point where the nasal tract splits into two choanae on each side of the septum wall. Each tube is represented by a length on the x axis and an area on the y axis.

move upwards. The constriction area varies from 0.8 cm^2 for a velum maximally open to 0 at occlusion, while its location, considered as the midpoint of the constriction tube, moves upwards simultaneously over a range of 0.9 cm. These observations are consistent with those of Demolin *et al.* (2003) who found for a velum maximally lowered a maximum of the constriction area ranging from 0.66 to 1.93 cm^2, depending on the subject. Moreover, they found that the cross-sectional area increases towards the cavum, as we observe in our data (see Fig. 17). The second parameter VS (Fig. 17(b)) induces smaller variations of the area function, but the relative variations of the constriction area might be large, up to a factor of 3. Contrary to the VL effects, the constriction area reduction controlled by VS is accompanied by a downward movement of the constriction location: while the constriction area decreases from 0.3 to 0.1 cm^2, its location moves about 0.2 cm towards the glottis.

C. Relation between velopharyngeal area function, velum height, and midsagittal distance

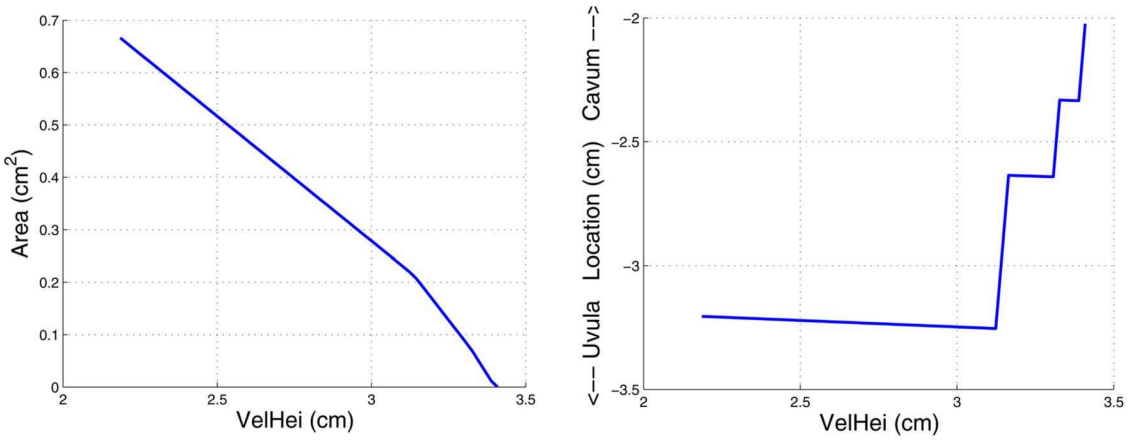
The present 3D articulatory model is very detailed, and thus rather complex to use. As access to 3D data and models is not commonly granted, we have derived from our data a simplified model that delivers directly the velopharyngeal area function as a function of velum height.

Velum height can be defined as the distance of a midsagittal velum point about halfway between the hard palate-velum junction and the tip of the uvula along an oblique line. In the present study, the oblique line obtained by linear regression on the 46 *pseudo-EMA vertex* of the data (Fig. 13(b)) is oriented at about 50° from an horizontal plane; velum height (referred to as *VelHei*) is thus defined as the distance along this line of the *pseudo-EMA vertex* projection from an arbitrary reference point (cf. the cross in Fig. 13(a)). The velopharyngeal port area function represents the velopharyngeal tract from the beginning of the nasal tract to the point where it splits into the two choanae. Considering that the tube number may change across articulations, the area function is resampled for each articulation into 11 tubes of identical length. The cross-sectional area of each of the 11 tubes is calculated so that the volume enclosed in it is equal to the total volume enclosed between the same boundaries along the midline in the original tract. The 11 cross-sectional areas and the common length of the tubes of the velopharyngeal area function are then linearly controlled by the *VelHei* parameter, the corresponding coefficients being determined by linear regression analysis over the set of 46 area functions.

As expected, a significant amount of the variance of the 11 tubes of the velopharyngeal port area function is explained by this parameter (87%). The constriction, which constitutes a critical parameter from the acoustic point of view, displayed an area RMS reconstruction error of 0.14 cm^2 and a position error of 0.41 cm. These errors are very close to those obtained when the articulatory model is controlled by only the VL parameter as in Sec. III G (area error of 0.16 cm^2 and position error of 0.38 cm). The variations of the constriction area and position as a function of *VelHei* are illustrated in Fig. 18. The linear modeling implies moreover that for each tube, area and length vary linearly with *VelHei*. Note that the step-like behavior of the constriction location can be ascribed to changes of the constriction tube index when the velum moves; increasing the number of tubes would reduce the step size without altering the general behavior. The velopharyngeal port area function is thus fairly well predictable directly from the midsagittal measurement of a single EMA point without resorting to an articulatory model.

A complementary approach was used by Björk (1961): based on sagittal and transverse x-ray tomograms taken from ten subjects, he established a linear relation between the minimal velum/pharyngeal wall midsagittal distance and the transverse constriction area for distances greater than 0.2 cm.

We applied the same approach with the articulatory model. A set of midsagittal contours was obtained from the 3D velum and pharyngeal wall meshes for VL varying between the extreme values found from the data. Figure 19 displays the relation between the minimal velum/pharyngeal wall midsagittal distance and the constriction area. This relation is not perfectly linear, but close to Björk's results.



a) Constriction area b) Constriction location

FIG. 18. (Color online) Variations of the area (a) and location (b) of the velopharyngeal port constriction as a function of Velum Height (*VelHei*). The origin of the tract is arbitrarily chosen as the point where the nasal tract splits into two choanae on each side of the septum wall.

Note that the sagittal distance is null for small values of the constriction area, which is explained by a full midsagittal contact between velum and pharyngeal wall while the channels on both sides remain open.

D. Influence of velum movement on the oral area function

An important question in the articulatory-acoustic study of nasals lies in the co-variation of the oral and nasal areas when the velum moves. Some researchers have assumed that the velum acts as a trap door (Fujimura and Lindqvist, 1971), while others have considered that the oral area decreases as much as the nasal area increases when the velum lowers, the total of both areas remaining constant (Maeda, 1982; Feng and Castelli, 1996). Feng and Kotenkoff (2006) have moreover suggested that the decrease of oral tract constriction area induced by the velum lowering may play an important role in nasalization. This present model allows studying this phenomenon.

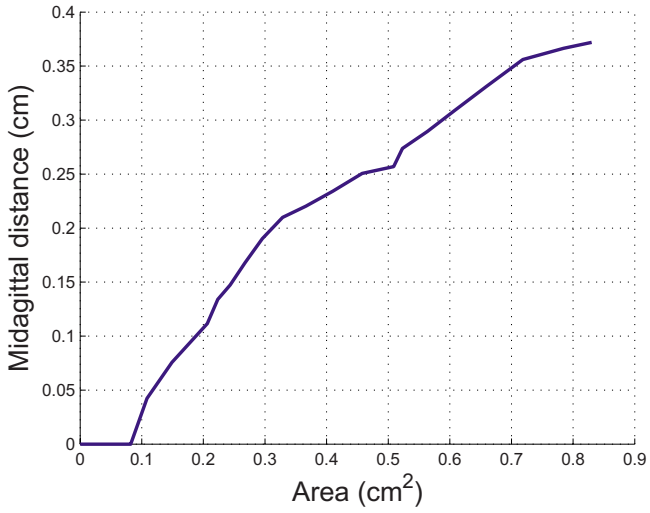


FIG. 19. (Color online) Variations of the velopharyngeal port midsagittal constriction distance as a function of the constriction area for VL varying linearly between the maximal values found in the data.

The influence of velum movements on the oral area function is illustrated by Fig. 20, which displays nomograms of the oral area function for the articulation [a] for VL and VS variations corresponding to those of Figs. 14 and 15. Note that the complete tract is considered fixed and that the area variations are only due to velum position variations as function of VL and VS. We observe, for this articulation, an important variation of the oral area function, from 0.2 to 3 cm^2 for the tube with maximal variation range between the two extreme positions of the velum corresponding to VL in the corpus, and from 1.4 to 2.3 cm^2 for the positions corresponding to VS. Important variations were observed for [i] and [u] as well, with areas varying from 5.1 to 8.5 cm^2 for [i] and from 0 to 2.9 cm^2 for [u] as a function of VL, and from 6.5 to 8 cm^2 for [i] and from 2.4 to 3.6 cm^2 for [u] as a function of VS.

In addition, we have characterized the co-variation of the areas of both tracts when the velum moves. Considering the relatively small variation of the oral area function with VS, we only investigated the effects of VL. As an illustration, Fig. 21 displays the variations of the oral and nasal areas of the first tube of the area functions at the location of the nasal tract branching for the three oral articulations [a], [i] and [u]. We observe that the oral area clearly increases faster than the nasal area decreases when the velum rises. The same behavior was found for the other tubes downstream, i.e., towards the lips or nostrils, though this asymmetric effect becomes less prominent.

V. INFLUENCE OF VELUM MOVEMENTS ON THE ACOUSTIC VOWEL SPACE

Nasalization is envisaged by Feng and Castelli (1996) as a trend from a pure oral articulation—from glottis to lips—to a pure nasopharyngeal target—from glottis to nostrils. These two extreme articulations correspond to single tracts without any other parallel tract (we do not consider in this study any of the various paranasal sinuses). The present model allows simulating the variations of the vocal tract induced by the velum alone, for transitions from a pure oral articulation to a

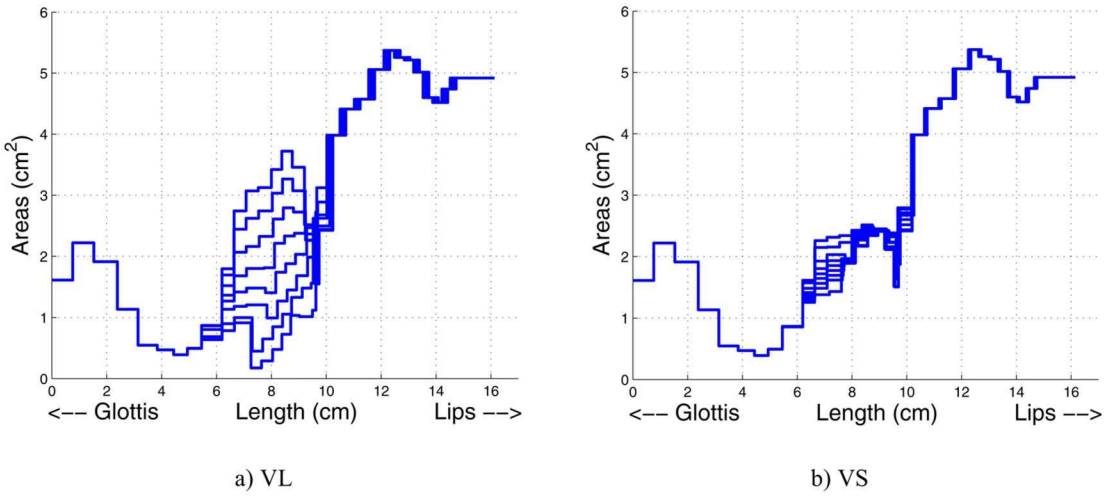


FIG. 20. (Color online) Variations of oral tract area function as a function of VL (a) and VS (b) parameters around the fixed articulation of [a]. The origin of the abscissa along the tract is arbitrarily chosen as glottal exit. Each tube is represented by a length on the x axis and an area on the y axis.

pure nasopharyngeal articulation, without influence of the coarticulatory movements usually observed on other organs, such as the tongue, in natural productions. The area functions of both tracts were determined according to the process described above. Using an acoustic model fed by these area functions, it is thus possible to investigate the influence of the velum movements on the acoustic transfer functions of the tracts between glottis and lips/nostrils. In the present study, we used a standard frequency domain electrical line analog model of acoustic plane wave propagation based on Badin and Fant (1984), and on Fant (1985) for the coupling of the nasal tract in parallel to the oral one. One may define at least three acoustic transfer functions: (1) the *oral transfer function*, i.e., the ratio between the acoustic velocity at the lips over the acoustic velocity at the glottis; (2) the *nasopharyngeal transfer function*, i.e., the ratio between the acoustic velocity at the nostrils over the acoustic velocity at the glottis; (3) the *total transfer function*, i.e., the sum of both, assuming as a first approximation that the sound radiated at some distance from the mouth and nostrils is just the sum of both contributions. The connection of a tract in parallel with the other tract both changes the poles of the transfer function of the tract considered and introduces zeros as well (see, e.g., Fant, 1985). Figure 22 displays, for example, the evolution of the *total* transfer function induced by variations of VL from a pure oral tract to a pure nasopharyngeal tract for a fixed articulation [a]. The two extreme *total* transfer functions, displayed in bold, correspond to the two articulatory targets, and contain only poles, while intermediate transfer functions contain poles and zeros due to the presence of parallel tracts.

We observe in the 500–1500 Hz frequency band that a bound pole-zero pair evolves from the first formant toward the second formant. Our simulations showed that the F1-F2

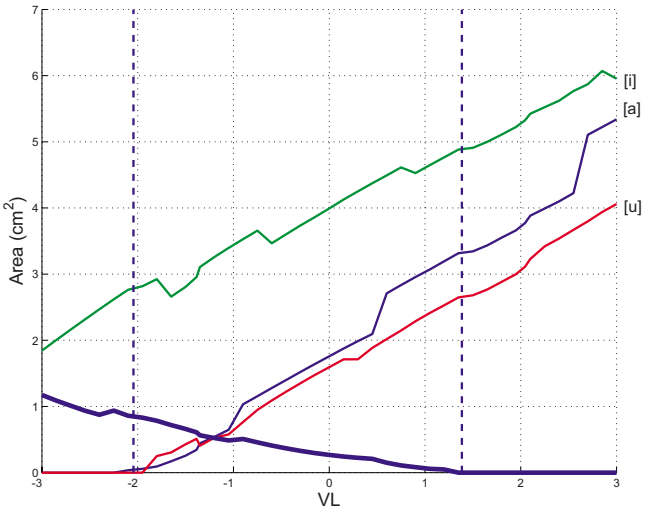


FIG. 21. (Color online) Oral (rising lines) and nasal (falling line) areas of the first tube of the area functions at the location of the nasal tract branching as a function of VL for the three articulations [a], [i] and [u]. The vertical dashed lines mark the minimal and maximal VL values found in the data.

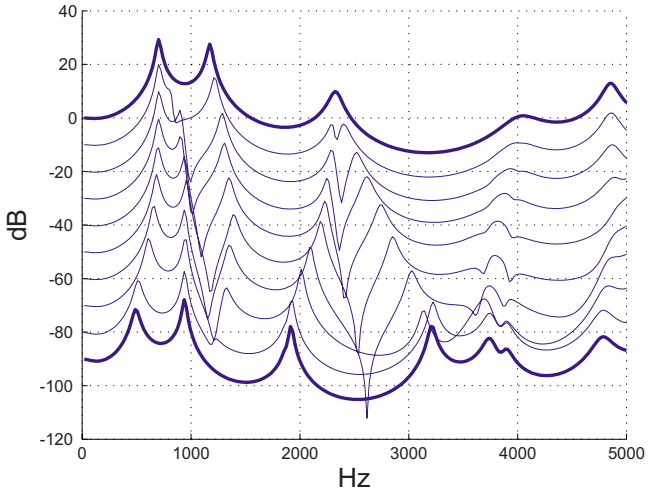


FIG. 22. (Color online) Nomograms of the *total* transfer function of the vocal tract obtained by varying VL from a pure oral tract (VL=1.4, upper bold line) to a pure nasopharyngeal tract (VL=−2.6, lower bold line) for a fixed articulation [a]. Each curve is arbitrarily separated from its neighbors by 10 dB for better reading. The two extreme transfer functions in bold correspond to oral tract (top) and nasopharyngeal tract (bottom) and contain only poles without zeros.

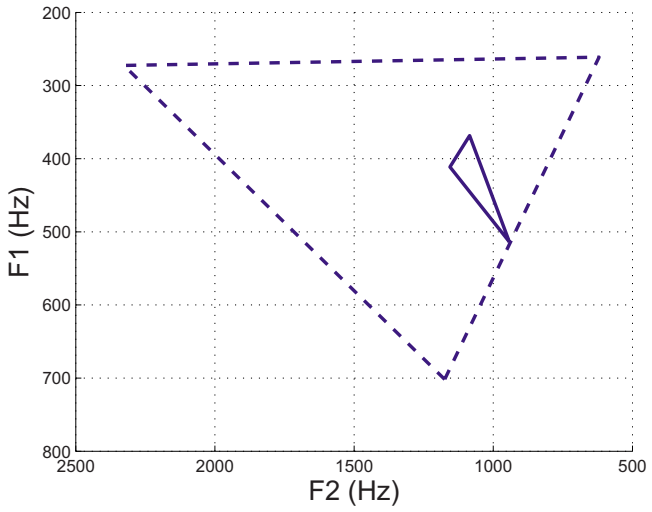


FIG. 23. (Color online) Representation in the acoustic F1-F2 vowel space of the pure oral articulations (high velum) for [a i u] (dashed lines) and of the corresponding pure nasopharyngeal articulations (low velum, solid lines).

formants associated with the pure nasopharyngeal tract for [a] are both lower than those associated with the pure [a] oral tract. Similar simulations for [i] and [u] led us to plot the vocalic triangle based on the three point vowels [a i u] for the two extreme positions of the velum (Fig. 23). The dashed triangle corresponds to the velum maximally high (pure oral tract) while the solid one corresponds to the velum maximally low (pure nasopharyngeal tract). We observe for the nasopharyngeal target a reduction of the vocalic space in a zone close to 450–1000 Hz, globally consistent with the simulations of Feng and Castelli (1996).

VI. SUMMARY AND PERSPECTIVES

A. Summary

This study has resulted in an original set of three-dimensional geometric data and in an original three-dimensional model of the velopharyngeal port, i.e., of the velum and of the nasopharyngeal wall. More precisely, sets of CT and MRI images have been collected from a French subject who sustained a set of 46 articulations covering the range of sounds that he can produce. Generic triangular surface meshes have been derived for the velum and the nasopharyngeal wall; their shapes, fitted by elastic deformation to each of the articulations, constitute accurate and consistent 3D geometric representations of these organs, with a constant number of vertices; they establish a unique reference collection of accurate 3D articulatory data. Moreover, each point of the generic meshes can be assumed to be a flesh point whose variations over the corpus represent the variability of a real flesh point. In addition, 3D shapes of rigid structures such as the nasal passages, the various paranasal sinuses, the hard palate, the jaw, etc. have been extracted from CT and MRI images of the subject. These data complement the various data already collected on the same French subject (cf. Beautemps *et al.*, 2001; Badin *et al.*, 2002; and Badin and Serrurier, 2006) and constitute a rich and accurate basis for further developments in articulatory and acoustic modeling of nasals and more generally of all articulations.

The second series of valuable results is the determination of the degrees of freedom of the velum and of the nasopharyngeal wall, and the associated 3D linear articulatory models. The first component extracted by PCA from the data was a head tilt movement in a sagittal plane; as it was not related to speech its contribution was removed from data. It was nevertheless hypothesized that this general head movement could induce an active compensation of the velum shape by the subject. A PCA applied to the velum data corrected for head tilt uncovered two main degrees of freedom: the first—dominant—component, that accounts for 83% of the velum coordinates' variance, corresponds to an oblique vertical movement clearly related to the pulling action that could be expected from the *levator veli palatini* muscle; the second component, that explains another 6% of the variance and controls smaller movements, represents a mostly horizontal movement that corresponds rather well to the more sphincter-like action plausible for the *superior pharyngeal constrictor*. It was also shown that the nasopharyngeal wall, though its variance is four times smaller than that of the velum, is very much correlated to the first parameter which was found to be able to account for 47% of its variance.

As a means to further assess the validity of the data and of the model, the midsagittal coordinates of a flesh point on the lower face of the velum were collected by means of an EMA system for the same subject uttering a comprehensive set of VCV sequences. It was found that the midsagittal space covered by this point is very similar to the midsagittal space covered by the particular point of the 3D velum generic mesh considered to be the same flesh point.

The velopharyngeal port model is thus efficiently controlled by these two articulatory parameters. Note that the RMS reconstruction error of the organs is less than one millimeter, which is consistent with results obtained on the tongue (cf. Badin *et al.*, 2002, and Badin and Serrurier, 2006).

Area functions could be derived from these 3D geometric representations. The precise 3D geometry of the nasal passages and the corresponding area functions constitute a new set of data consistent with the few investigations found in the literature and confirms in particular a constriction of about 1 cm^2 large and 0.6 cm long located 1.7 cm before nostrils outlet. The nasal coupling area was found to raise up to 0.8 cm^2 while the corresponding constriction location may move over a range of about 1 cm. In addition, a simplified model of velopharyngeal port area function controlled directly by the velum height was derived. The strong influence of the velum movement on the oral tract area function has been confirmed. Finally, acoustic simulations based on the model gave results similar to that of Feng and Castelli (1996), i.e., a reduction of the F1-F2 acoustic vowel space when going from pure oral vowels to the corresponding pure nasopharyngeal tract in the 450–1000 Hz zone, sometimes referred to as the *nasal gap*.

B. Perspectives

A part of the complex geometry of the nasal passages remained unexploited in the present study. The shape of the various sinuses should be integrated into the area functions of the nasal passages, and their influence on the acoustics should be further assessed. More generally, the formants determined from the articulatory model should be compared to those measured on the subject for the same articulations in order to validate the model at the acoustic level. In particular, separate mouth-nose acoustic recordings performed on the same subject by Feng and Kotenkoff (2006) will reveal very useful information. It is also likely that the complex shape of the nasal passages should be used to elaborate more realistic acoustic models, including more realistic losses.

In order to be fully operational for the study of nasality, the new model of velopharyngeal port presented should be integrated into a complete articulatory model, including especially the tongue (cf. Badin and Serrurier, 2006). This would allow accurate control of the complete 3D geometry of the oral and nasal tracts, to obtain corresponding area functions from the glottis to the lips and the nostrils, and finally to determine the resulting acoustic characteristics in terms of formants or speech sounds. This complete model also opens the possibility to characterize the importance of tongue backing for French nasal vowels as discussed by Maeda (1993), Delvaux *et al.* (2002), or Feng and Kotenkoff (2006). The complete model would also allow the determination of the covariation between oral and nasal coupling areas related to velum movements usually assumed to be linear in the literature (see, e.g., Maeda, 1982, or Feng and Castelli, 1996).

The use of this articulatory model makes it possible to study the influence of sole velum movements on the vocal tract without the coarticulatory movements usually made by other organs during normal velum transitions in natural speech. More precisely, in the continuity of the work led by Feng and Castelli (1996) and Feng and Kotenkoff (2006), it will be useful to perform simulations to assess the acoustic consequences of the velum movement separately on the velopharyngeal constriction, oral constriction and nasal coupling place in the oral tract, in order to contribute to the characterization of the acoustic cues of nasality.

The model, in relation with the dynamic articulatory midsagittal trajectories of fleshpoints on the velum and the tongue provided by the EMA device, will allow the *articulatory copy synthesis* of VCV sequences through the recovery of the complete 3D geometry of the vocal tract. The control of the various articulators for dynamic speech could thus be investigated. In particular, assuming that the principal parameter of the velopharyngeal port, e.g. VL, allows the control of the constriction from closure to maximal aperture, it would be interesting to evaluate the possibility of controlling the velum with only the VL parameter, which would reduce the complexity of articulatory control.

Finally, the articulatory synthesizer available would be useful to study the perception of nasality. In particular, the study of the interaction between nasalization and height in vowel perception led by Macmillan, Kingston, Thorburn, Walsh Dickey, and Bartels (1999) using a formant synthesizer could be replicated with more ecological stimuli, i.e., stimuli that could really be produced by a human subject.

ACKNOWLEDGMENTS

We would like to thank Christoph Segebarth, Jean-François Lebas, and Monica Baciu for the main corpus sagittal MR images acquisition at the Radiology Department of the University Hospital, Grenoble, France, and Kiyoshi Honda, Shinobu Masaki, Sayoko Takano, Yasuhiro Shimada, and Ichiro Fujimoto for coronal MRI and head tilt MRI acquisition at ATR, Kyoto, Japan. The EMA data have been recorded at ICP, Grenoble, France, thanks to the help of Alain Arnal, Christophe Savariaux, and Solange Rossato. We are grateful to Jacques Lebeau and Franz Chouly for their valuable discussions and bibliographic advice about anatomy. We acknowledge the help of the GMCAO team at TIMC for the use of the elastic matching software *TestRigid* (Yohan Payan, Maxime Bérar), Daniel Huber for the smoothing software *SmoothMesh* (VMR Lab at Carnegie Mellon University, USA) and Andreas Fabri (Geometrica Research Group at INRIA, France) for the 3D meshing software. We are also grateful to David Pritchard for the English corrections. Finally, we would like to thank three anonymous reviewers and the associate editor of the journal, Christine Shadle, for their pertinent comments on the first and second versions of the manuscript, and for their very detailed and helpful editorial help.

Amelot, A., Crevier-Buchman, L., and Maeda, S. (**2003**). "Observations of the velopharyngeal closure mechanism in horizontal and lateral directions from fiberscopic data," In *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M.-J. Solé, D. Recasens, and J. Romero, Barcelona, Spain, pp. 3021–3024.

Badin, P., Bailly, G., Elisei, F., and Odisio, M. (**2003**). "Virtual Talking Heads and Audiovisual Articulatory Synthesis (Invited talk at the symposium "Articulatory synthesis. Advances and prospects")." In *Proceedings of the 15th International Congress of Phonetic Sciences*, M.-J. Solé, D. Recasens, and J. Romero, editors, Barcelona, Spain, Vol. **1**, pp. 193–197.

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C., and Savariaux, C. (**2002**). "Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images," J. Phonetics **30**(3), 533–553.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computation," Speech Transmission Laboratory-Quarterly Progress Status Report-Stockholm, **25**, 53–108.

Badin, P., Motoki, K., Miki, N., Ritterhaus, D., and Lallouache, T. M. (**1994**). "Some geometric and acoustic properties of the lip horn," J. Acoust. Soc. Jpn. (E) **15**(4), 243–253.

Badin, P., and Serrurier, A. (**2006**). "Three-dimensional modeling of speech organs: Articulatory data and models," In *Transactions on Technical Committee of Psychological and Physiological Acoustics*, The Acoustic Society of Japan, Kanazawa, Japan, Vol. **36**(5), H-2006-77, pp. 421–426.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am. **90**(2, Pt. 1), 799–828.

Beautemps, D., Badin, P., and Bailly, G. (**2001**). "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling," J. Acoust. Soc. Am. **109**(5), 2165–2180.

Bell-Berti, F. (**1976**). "An electromyographic study of velopharyngeal function in speech," J. Speech Hear. Res. **19**, 225–240.

Bell-Berti, F. (**1993**). "Understanding velic motor control: Studies of segmental context," In *Phonetics and Phonology, Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow, (Academic, New York), pp. 63–85.

Björk, L. (**1961**). *Velopharyngeal Function in Connected Speech. Studies Using Tomography and Cineradiography Synchronized with Speech Spec-*

trography (Appelbergs Bocktryckeri AB, Stockholm, Sweden).
Bjuggren, G., and Fant, G. (**1964**). "The nasal cavity structures," STL-QPSR **5**(4), 5–7.
Couteau, B., Payan, Y., and Lavallée, S. (**2000**). "The mesh-matching algorithm: An automatic 3D mesh generator for finite element structures," J. Biomech. **33**, 1005–1009.
Crystal, D. (**1997**). *A Dictionary of Linguistics and Phonetics* (Blackwell, MA).
Dang, J., Honda, K., and Suzuki, H. (**1994**). "Morphological and acoustical analysis of the nasal and the paranasal cavities," J. Acoust. Soc. Am. **96**, 2088–2100.
Delvaux, V., Metens, T., and Soquet, A. (**2002**). "French nasal vowels: Acoustic and articulatory properties," In *Proceedings of the 7th International Conference on Spoken Language Processing*, Denver, Vol. **1**, pp. 53–56.
Demolin, D., Delvaux, V., Metens, T., and Soquet, A. (**2003**). "Determination of velum opening for French nasal vowels by magnetic resonance imaging," J. Voice **17**(4), 454–467.
Demolin, D., Lecuit, V., Metens, T., Nazarian, B., and Soquet, A. (**1998**). "Magnetic resonance measurements of the velum port opening," In *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by R. H. Mannell and J. Robert-Ribes, Sydney, Australia, Australian Speech Science and Technology Association Inc., Vol. **2**, pp. 425–428.
Dickson, D. R., and Dickson, W. M. (**1972**). "Velopharyngeal anatomy," J. Speech Hear. Res. **15**, 372–381.
Engwall, O., and Badin, P. (**1999**). "Collecting and analyzing two- and three-dimensional MRI data for Swedish," Tal Musik Hörsel - Quarterly Progress Status Report - Stockholm, **40**(3–4), 11–38.
Fabri, A. 3D meshing software (http://cgal.inria.fr/Reconstruction, link available **May 29, 2007**).
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Fant, G. (**1985**). "The vocal tract in your pocket calculator," STL-QPSR **26**(2–3), 1–19.
Feng, G., and Castelli, E. (**1996**). "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization," J. Acoust. Soc. Am. **99**(6), 3694–3706.
Feng, G., and Kotenkoff, C. (**2006**). "New considerations for vowel nasalization based on separate mouth-nose recording," In *Proceedings of Interspeech'06*, Pittsburgh, pp. 2242–2245.
Ferguson, C. A., Hyman, L. M., and Ohala, J. J. (Editors), (**1975**). *Nasalfest: Papers from a symposium on nasals and nasalization*, Stanford, CA.
Fujimura, O., and Lindqvist, J. (**1971**). "Sweep-Tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. **49**, 541–558.
House, A. S., and Stevens, K. N. (**1956**). "Analog studies of the nasalization of vowels," J. Speech Hear Disord. **21**, 218–232.
Huber, D., The SmoothMesh software (http://www.cs.cmu.edu/~vmr/software/meshtoolbox/downloads.html, link available at **May 29th, 2007**).
Huffman, M. K., and Krakow, R. A. (**1993**). *Phonetics and Phonology. Nasals, Nasalization, and the Velum* (Academic Press, New York), Vol. **5**.
Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (**1986**). "The dynamical theory of speech production: Data and theory," J. Phonetics **14**, 29–60.
Kent, R. D. (**1997**). *The Speech Sciences* (Singular, Stockholm, Sweden).
Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N., and 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.
Kollia, H. B., Gracco, V. L., and Harris, K. S. (**1995**). "Articulatory organization of mandibular, labial, and velar movements during speech," J. Acoust. Soc. Am. **98**(3), 1313–1324.
Macmillan, N. A., Kingston, J., Thorburn, R., Walsh Dickey, L., and Bartels, C. (**1999**). "Integrality of nasalization and F_1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions," J. Acoust. Soc. Am. **106**(5), 2913–2932.
Maeda, S. (**1982**). "The role of the sinus cavities in the production of nasal vowels," In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Paris, France, pp. 911–914.
Maeda, S. (**1993**). "Acoustics of vowel nasalization and articulatory shifts in French nasal vowels," In *Phonetics and Phonology. Nasals, nasalization and the velum* (Academic, New York), pp. 147–167.
Matsumura, M., Niikawa, T., Shimizu, K., Hashimoto, Y., and Morita, T. (**1994**). "Measurement of 3D shapes of vocal tract, dental crown and nasal cavity using MRI: Vowels and fricatives," In *Proceedings of the 3rd International Conference on Spoken Language Processing*, Yokohama, Japan, Vol. **2**, pp. 619–622.
Matsumura, M., and Sugiura, A. (**1992**). "Modeling of 3-dimensional vocal tract shapes obtained by magnetic resonance imaging for speech synthesis," In *Proceedings of the 2nd International Conference on Spoken Language Processing*, Banff, Canada, pp. 425–428.
Mermelstein, P. (**1973**). "Articulatory model for study of speech production," J. Acoust. Soc. Am. **53**, 1070–1082.
Rossato, S., Badin, P., and Bouaouni, F. (**2003**). "Velar movements in French: An articulatory and acoustical analysis of coarticulation," In *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M.-J. Solé, D. Recasens, and J. Romero, Barcelona, Spain, pp. 3141–3144.
Serrurier, A., and Badin, P. (**2005a**). "A three-dimensional linear articulatory model of velum based on MRI data," In *Interspeech'2005 - Eurospeech - 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, pp. 2161–2164.
Serrurier, A., and Badin, P. (**2005b**). "Towards a 3D articulatory model of velum based on MRI and CT images," ZAS Papers in Linguistics, Speech production and perception: Experimental analyses and models, edited by Susanne Fuchs, Pascal Perrier and Bernd Pompino-Marschall, Vol. 40, pp. 195–211.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from Magnetic Resonance Imaging," J. Acoust. Soc. Am. **100**(1), 537–554.
Sundberg, J., Johansson, C., Wilbr, H., and Ytterbergh, C. (**1987**). "From sagittal distance to area. A study of transverse, vocal tract cross-sectional area," Phonetica **44**, 76–90.
Takemoto, H., Kitamura, T., Nishimoto, H., and Honda, K. (**2004**). "A method of teeth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions," Acoust. Sci. & Tech. **25**, 468–474.
Teixeira, A., Moutinho, L. C., and Coimbra, R. L. (**2003**). "Production, acoustic and perceptual studies on European Portuguese nasal vowels height," In *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M.-J. Solé, D. Recasens, and J. Romero, Barcelona, Spain, pp. 3033–3036.
Teixeira, A., Vaz, F., Moutinho, L., and Coimbra, R. L. (**2001**). "Articulory Synthesis of Portuguese," In *III Encontro do Forum International de Investigadores Portugueses* (Institut EETA, Aveiro, Portugal).
Tiede, M. K., Masaki, S., and Vatikiotis-Bateson, E. (**2000**). "Contrasts in speech articulation observed in sitting and supine conditions," In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, Kloster Seeon, Germany, pp. 25–28.
Wrench, A. A. (**1999**). "An investigation of sagittal velar movement and its correlation with lip, tongue and jaw movement," In *Proceedings of the 14th International Congress of Phonetic Sciences* edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, San Francisco, pp. 2259–2262.
Zemlin, W. R. (**1968**). *Speech and Hearing Science-Anatomy and Physiology* (Prentice-Hall, NJ).