Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling

Denis Beautemps,[a)] Pierre Badin,[b)] and Gérard Bailly[c)]
Institut de la Communication Parlée, UMR CNRS 5009, INPG–Université Stendhal, 46, Av. Félix Viallet, F-38031 Grenoble Cedex 1, France

(Received 15 August 1999; accepted for publication 8 February 2001)

The following contribution addresses several issues concerning speech degrees of freedom in French oral vowels, stop, and fricative consonants based on an analysis of tongue and lip shapes extracted from cineradio- and labio-films. The midsagittal tongue shapes have been submitted to a linear decomposition where some of the loading factors were selected such as jaw and larynx position while four other components were derived from principal component analysis (PCA). For the lips, in addition to the more traditional protrusion and opening components, a supplementary component was extracted to explain the upward movement of both the upper and lower lips in [v] production. A linear articulatory model was developed; the six tongue degrees of freedom were used as the articulatory control parameters of the midsagittal tongue contours and explained 96% of the tongue data variance. These control parameters were also used to specify the frontal lip width dimension derived from the labio-film front views. Finally, this model was complemented by a conversion model going from the midsagittal to the area function, based on a fitting of the midsagittal distances and the formant frequencies for both vowels and consonants. [DOI: 10.1121/1.1361090]

PACS numbers: 43.70.Bk [AL]

I. INTRODUCTION

''Speech is rather a set of movements made audible than a set of sounds produced by movement,'' posited Stetson in 1928 (p. 29). This statement could be more properly rephrased as ''Speech can be regarded as the *audible* and *visible* signals resulting from articulatory movement,'' as stated, for instance, in the *speech robotics* approach fostered in the collaborative European project *Speech Maps* (Abry *et al.*, 1994). In this framework, the speech apparatus is viewed as a *plant* driven by a *controller* so as to recruit articulators and coordinate their movements, which have simultaneous acoustic and visual consequences.

The concept of plant and controller implies the notion of a relatively small number of *independent degrees of freedom* for the articulatory plant, i.e., the specification, for each articulator, of a limited set of movements that can be executed independently of each other by the articulator. As emphasized by Kelso *et al.* (1986), however, the speech production apparatus is made of a large number of neuromuscular components that offer a potentially huge dimensionality and which must be functionally coupled in order to produce relatively simple gestures [this view forms the basis of the concept of coordinative structures in speech, cf. Fowler and Saltzman (1993)]. Maeda (1991) refers to a similar concept in terms of ''elementary articulators.''

One *independent degree of freedom* may be more precisely defined for a given speech articulator as one variable that can completely control a specific variation of shape and position of this articulator, and that is statistically independent of the other degrees of freedom over a 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 mainly three levels of implicit or explicit constraints: (1) physical continuity of the articulators (the tongue cannot have a jigsaw shape for instance); (2) biomechanical constraints (the range of possible articulators shapes and positions is limited by the physiological properties of the bony structures and of the muscles); and (3) the nature of the task in relation with control (chewing involves lateral translations of the jaw, but speech does not, and thus jaw has different degrees of freedom depending on the task observed). The correlations observed on articulatory measurements thus cannot always be ascribed with certainty to either biomechanical constraints or to strategies related to the task. For instance, Hoole and Kroos (1998) observed that larynx height and lip protrusion are inversely correlated: this correlation obviously cannot be explained by biomechanical links between lips and larynx, but should be ascribed to control strategies related to the speech task. It thus appears understandable that determining which properties of speech can be attributed to the plant and which to the controller is a recurrent issue in speech motor control (cf., e.g., Perkell, 1991; Scully, 1991; Abry *et al.*, 1994).

It has long been known that midsagittal profiles constitute a privileged representation of speech articulation (cf., e.g., Boë *et al.*, 1995, for a review on vowel representations). Indeed, for most phonemes, the complete vocal tract shape can be fairly well inferred from the midsagittal plane, the most notable exception being lateral sounds. Moreover, midsagittal profiles allow linking of vocal tract articulation and

[a)]Electronic mail: beautemps@icp.inpg.fr
[b)]Electronic mail: badin@icp.inpg.fr
[c)]Electronic mail: bailly@icp.inpg.fr

the resulting acoustics. Articulatory models can therefore be viewed as one of the most efficient means of manipulating vocal tract shapes, and midsagittal profiles as a privileged interface between *motor control* on the one hand and *acoustic and visual* modules of the speech production system on the other hand. Developing and evaluating such articulatory models finally constitutes a good means for identifying the degrees of freedom of speech articulators.

Among the large number of studies devoted to articulatory modeling since the seventies, two main approaches can be identified: *functional articulatory modeling*, where the position and shape of articulators are algebraic functions of a small number of articulatory parameters, and explicit *biomechanical modeling*, where the position and shape of articulators are computed from physical simulations of the forces generated by muscles and of their consequences on the articulators.

In linear articulatory models, the relations between articulator positions and shapes and the control parameters can either be defined in geometrical terms, in which case the degrees of freedom of the articulatory plant are decided *a priori* and fitted to the data *a posteriori* (cf., e.g., Coker and Fujimura, 1966; Liljencrants, 1971; Mermelstein, 1973), or based on articulatory data measured on one or several subjects, in which case the degrees of freedom of the plant emerge from the data (cf., e.g., Lindblom and Sundberg, 1971; Maeda, 1990; Stark *et al.*, 1996).

The general approach of biomechanical articulatory models consists in modeling muscular forces and articulator structure by means of methods inspired from mechanical analysis and numerical simulation (cf., e.g., Perkell, 1974; Wilhelms-Tricarico, 1995; Laboissière *et al.*, 1996; Payan and Perrier, 1997). These models present the advantage of being physical models with intrinsic dynamics, although necessarily extremely simplified, but their control remains very complex, in particular due to the high number of degrees of freedom represented by each individual muscle command. Sanguineti *et al.*'s (1998) work constitutes a good illustration of this. They fitted, with their model, the articulator shapes and positions measured from the x-ray database already used by Maeda (1990) and determined, by optimization, the commands of the 17 muscles involved in their model. They identified then, by linear component analysis applied in the so-called λ-space corresponding to the biomechanical tongue control parameter space, the synergies between these commands, and showed that six independent components could account for most of the data variance of the midsagittal tongue shape. These first six components are closely related to the degrees of freedom that could be extracted directly from the original x-ray contours (Maeda, 1990). It thus appears that, from the point of view of the degrees of freedom, such complex biomechanical models are not a prerequisite to the accurate description of static speech articulation.

Rather than developing *a priori* complex biomechanical models with degrees of freedom in large excess and then reducing this high dimensionality based on articulatory data, we have adopted a dual approach in the present work. More precisely, our objectives were to determine the linear degrees of freedom of one subject's articulators in the midsagittal plane, and to build an *articulatory-acoustic plant* that could be considered a faithful and coherent representation of this subject's articulatory and acoustic capabilities.

The present paper describes our approach to this problem: (1) design of the corpus and collection of articulatory-acoustic data for one subject, (2) analysis of the data and extraction of the independent linear degrees of freedom of the articulators, and (3) the development of a linear articulatory-acoustic model based on these degrees of freedom.

II. ARTICULATORY AND ACOUSTIC DATA ACQUISITION METHODOLOGY

A. The experimental setup: Synchronized cineradio- and labio-film

Cineradiography was chosen as the best compromise between good spatial and temporal resolutions for the articulatory data. This technique, which has been used successfully for speech studies at the Strasbourg Phonetic Institute (cf., e.g., Bothorel *et al.*, 1986), was used in synchrony with the video labiometric method developed at ICP by Lallouache (1990) (cf. also Badin *et al.* 1994a). The recordings were performed at the Strasbourg Schiltigheim Hospital, France. The subject's head was positioned at a distance of 50 cm from the x-ray emitter and 20 cm from the radiance amplifier. An aluminum filter was placed in the lip region to avoid overexposure of the lips, thus improving the contrasts in this region (cf. Bothorel *et al.*, 1986). The vocal tract images produced by the radiance amplifier were captured and recorded by a 35-mm film camera. The subject's lips, painted in blue to allow the lip contours to be extracted by an image processing procedure, were recorded using a video camera. Both cameras were operating at a rate of 50 frames per second. The speech signal, captured by a directional microphone placed at a distance of 10 cm from the subject's mouth, was synchronously recorded.

B. The subject

The choice of the set of subjects always poses a dilemma: a single subject study surely reduces the generality of the work but allows us to gather rich and detailed data, whereas a study with a larger panel of subjects may permit us to draw some general conclusions but limits the extent of the data that can be practically acquired and processed.

Under the auspices of the European collaborative project *Speech Maps*, Abry *et al.* (1994) aimed to gather a variety of converging and complementary articulatory/acoustic data for one subject uttering the same speech material in a controlled manner in different experimental setups. This policy resulted in a large set of data of potential use in speech production modeling, such as vocal tract acoustic transfer functions (Djéradi *et al.*, 1991), acoustic and aerodynamic pressure and flow in the tract (Stromberg *et al.*, 1994; Badin *et al.*, 1995; Shadle and Scully, 1995), electropalatography data (Badin *et al.*, 1994b), and more recently 3D MRI vocal tract images (Badin *et al.*, 1998, 2000) and video face data (Badin

TABLE I. Corpus used for the cineradiographic recordings.

[aɛeiyuoø]
[pavapavipavupivipivupivy]
[pazapazipazupizipizupizy]
[paʒapaʒipaʒupiʒipiʒupiʒy]
[abaabiabuibiibuiby]
[adaadiaduidiiduidy]
[agaagiaguigiiguigy]

et al., 2000). As the present study was conducted in the framework of this project, the same subject was therefore chosen as a *reference* subject.

C. The corpus

As mentioned in the previous section, one of the major aims of the present study was to determine the degrees of freedom of a subject's articulators for speech, excluding any other nonspeech movements. Attaining this goal would ideally require recording a large corpus of speech material containing all possible combinations of phonemes. This is obviously not practical in general, and particularly inappropriate in the case of cineradiography, due to health hazards related to this method. The corpus was thus designed to include as many combinations of Vowel Consonant Vowel (VCV) sequences as possible in a very limited amount of time.

The voiced French plosive and fricative consonants C=[vzʒbdg] were chosen in six vocalic contexts involving the four French extreme vowels V=[aiuy]:aCa, aCi, aCu, iCi, iCu, iCy. It was assumed that the voiceless cognates of these consonants correspond approximately to the same articulation. The presence of voicing was expected to simplify the tracking of formants during fricative consonants. The French [l] was excluded because of the impossibility of getting information on the lateral channels from midsagittal x-ray pictures. Nasals were also excluded because velar movements do not directly influence other articulators movement (although nasal vowels in French seems to imply some additional tongue backing compared to other vowels, cf., e.g., Zerling, 1984), and will thus be studied in the near future. Finally the French [ʁ] was also not included because it was not hypothesized to require extra degrees of freedom for the midsagittal profile. The extreme vowels were expected to represent the most extreme vocalic articulations in French. Moreover, the fricative items were interspersed with [p]'s in order to allow the estimation of subglottal pressure during the fricatives as the intraoral pressure during the closure of [p] (Demolin *et al.*, 1997). In addition, a series of connected vowels [aɛeiyuoø] was recorded in order to test formant/cavity affiliation hypotheses (Bailly, 1993). Finally, the corpus duration could be reduced to about 24.5 s of signal (actually leading to 1222 pictures), with the following distribution of phonemes: 30 [ai], 12 [u], 6 [y], 6 [vzʒbdg], 18 [p]. Table I presents the complete corpus.

D. Processing of the x-ray and video images

For each picture, the sagittal contours were first drawn by hand from a projection of the picture onto a piece of paper, digitized by a scanner, and finally 11 subcontours were hand-edited by means of an interactive software. These subcontours correspond to different articulators or vocal tract regions: the upper and lower lips, the hard palate, the velum, the different components of the pharynx, the larynx, the tongue, the jaw, and the hyoid bone. Figure 1 presents an example of the resulting midsagittal contour. Note that rigid structures (hard palate, jaw, hyoid bone) were not drawn again for each image, but were given reference contours that best fitted most of the shapes observed on the whole set of images; for these structures, the operator's task was then only to optimally position the reference shapes for each image by roto-translation. This procedure presented four advantages: (1) the operator's task was made easier and faster; (2) it reduced the noise due to manual drawing; (3) it avoided the difficulty of precisely determining reference landmarks such as incisor edges from images where the contrast is not always very high; (4) it offered the possibility of determining in a straightforward manner the positions of these rigid bodies in the midsagittal plane (see the discussion on jaw analysis below).

Concerning the lips, the blue of the video front lip images was converted into absolute black by means of an analogue Kroma-key, and the inner contour was then automatically determined by simple adapted thresholding (cf. Lallouache, 1990, or Badin *et al.*, 1994a, for more details).

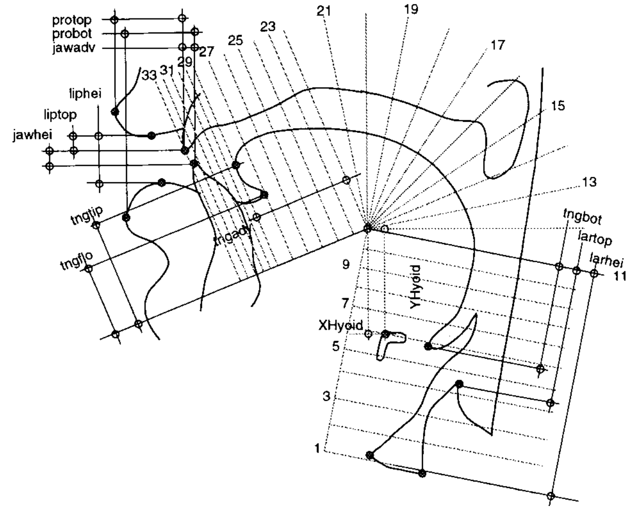

FIG. 1. Example of manually drawn *VT* contours and associated articulatory measures: upper lip protrusion *ProTop*, lower lip protrusion *ProBot*, upper lip elevation *LipTop*, jaw height *JawHei*, jaw advancement *JawAdv*, tongue tip advancement *TngAdv*, tongue tip height *TngTip*, tongue floor height *TngFlo*, tongue bottom *TngBot*, height of the larynx top *LarTop*, height of the larynx bottom *LarHei*, *x/y* coordinates of the hyoid bone *XHyoid,YHyoid*.

E. Articulatory measurements

Before going into some details, it is useful to specify the midsagittal coordinate system attached to the skull structure and used in this study. The lower edge of the upper incisors is given arbitrary *x/y* coordinates (5,10) in cm. The *x*-axis is positive in the *posterior* direction of the head (toward the back), and negative in the *anterior* direction (toward the nose); the *y*-axis is positive in the *superior* direction (toward the brain), and negative in the *inferior* direction (toward the feet). Finally, it happens that the direction of the *maxillary occlusal plane* (defined as the plane ''given by the tips of the central incisors and at least two other maxillary teeth on opposite sides of the mouth,'' Westbury, 1994), is oriented at an angle of 4.6° from the $y=0$ axis.

The values of a number of geometrical parameters (see Fig. 1) have been determined from the midsagittal contours: upper *ProTop* and lower *ProBot* lip protrusions, upper lip elevation *LipTop*, jaw height *JawHei*, jaw advancement *JawAdv*, tongue tip advancement *TngAdv* and height *TngTip*, tongue floor height *TngFlo*, tongue bottom *TngBot*, height of the larynx top *LarTop* and bottom *LarHei*, x/y coordinates of the hyoid bone *XHyoid*, *YHyoid*. In addition, three parameters were extracted from the video front views of the lips: lip height *B*, lip width *A*, and the intra-labial lip area *S*.

Note that the distance between the upper and lower lips can be determined either from the midsagittal profile *LipHei* or from the front view *B*: as expected, the two measures are very close to each other, *B* being less accurate when the lip opening is close to zero, particularly for the rounded vowels [uy], due to the fact that the subject's upper lip tends to mask the intra-labial orifice in such cases. The correlation coefficient between the measures is 0.99, while the rms error is 0.1 cm.

The articulators are expected to follow relatively smooth trajectories due to their long time responses (cf., e.g., the jaw characteristic resonance frequency of 5–6 Hz mentioned by Sorokin *et al.*, 1980). Deviations of geometric measures from their smooth trajectories revealed that the noise added by the whole chain of acquisition was in the range of 1-mm peak-to-peak.

F. Midsagittal contours

A semi-polar grid has been used to describe the midsagittal contours, as has traditionally been done since Heinz and Stevens (1965) or Maeda (1979). However, as proposed by Gabioud (1994), two parts of this grid have been made adjustable in order to follow the movements of the larynx (gridlines 1 to 6, line 1 being the lowest one near the glottis) and the movement of the tongue tip (in fact, the *tongue blade*, defined as the linguistic class *coronal articulation*; grid lines 24 to 28). This grid presents a double advantage: (1) the number of intersection points between the grid and the tongue contour is constant, whatever the extension of the tongue tip or of the larynx, which is a crucial feature for further statistical analysis: (2) the fact that the measurement grid follows tongue tip movements implies that all the points in the vicinity of the tongue tip present a behavior close to that of *flesh-points*, i.e., points mechanically attached to the tongue surface, which is more than just a description of tongue shape (this is useful when recovering tongue contours from flesh-points coordinates measured by electromagnetic articulometry; see Badin *et al.*, 1997). Finally, a third part of the grid has been introduced to describe the alveolar dental cavity with adjustable grid lines (grid lines 29 to 33) equally spaced between the tongue tip and the lower edge of the upper incisor. The inner and outer vocal tract midsagittal contours intersect thus the grid lines at 2×33 points; each contour can thus be represented by the 33-element vector (referred to as *Int* and *Ext*) of the abscissa of these intersection points along the grid lines. Note that, as mentioned by Westbury (1994), the dimensionality of the articulators may change depending on the coordinate system used. For instance, a point running on a fixed circle appears to have two *linearly* independent degrees of freedom in a Cartesian system, but only one single degree of freedom in a polar coordinate system. The choice of the dynamically adjustable semi-polar grid system seems a good solution to avoid artificial overdimensionality.

The velum is terminated in the midsagittal plane by the uvula. It can be in contact with the upper surface of the tongue, however, without creating a real constriction in the vocal tract, since air remains free to flow on each side. This fact has been approximately taken into account by making the velum artificially thinner by a factor linearly increasing from 0 at its extremity to about 40% at its base, and by shifting the result so as to align its posterior wall with the pharyngeal wall.

G. Formants

The speech signal was digitized at 16 kHz, and the first four formant values were estimated using LPC analysis with a 20-ms window centered on the times where the midsagittal views were acquired, leading to formant trajectories sampled at 50 Hz. Because of the background noise due to the x-ray emitter, the signal-to-noise ratio was rather poor (about 25 dB), and thus some of the formants had to be hand-edited. This was done in reference to another version of the same corpus recorded by the subject in good recording conditions. The $F1/F2$ and $F1/F3$ spaces for the pooled vowels and consonants are shown in Fig. 2.

III. ANALYSIS OF THE INDEPENDENT LINEAR DEGREES OF FREEDOM OF THE MIDSAGITTAL CONTOURS

A. Principles

1. Identifying degrees of freedom

As mentioned in Sec. I, the approach taken in the present study to determine the degrees of freedom of the various speech articulators is based on articulatory data obtained from one *subject* producing a given *corpus* in a given *language*.

In general, speech articulators possess excess degrees of freedom, i.e., a given articulation can be achieved by means of different combinations of the available degrees of freedom of the articulators (cf. bite-block experiments performed by Lindblom *et al.*, 1979). Control strategies finally aim at recruiting these degrees of freedom when they are needed to attain given articulatory/acoustic/visual goals, and leaving them free to anticipate other goals whenever possible (this is one basic principle of *coarticulation*). In the present data driven approach, the problem is to decide the repartition of the variance of the measured articulatory variables between the different variables associated with the degrees of freedom. The present work rests on a common consideration in speech motor control modeling: what is explained by the biomechanics of the speech plant does not need to be worked

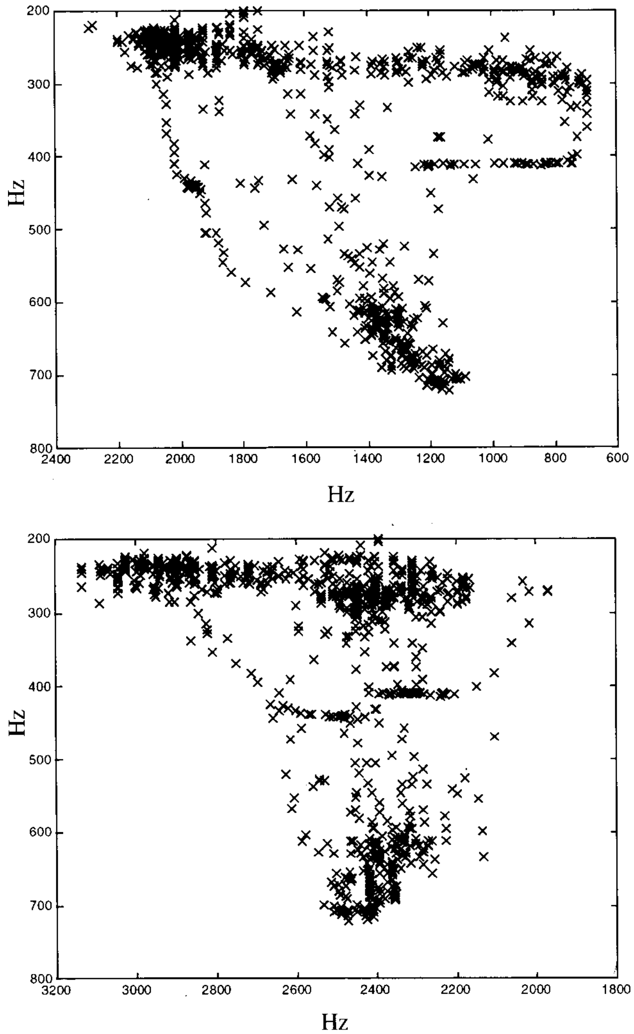

FIG. 2. Measured $F2/F1$ (top) and $F3/F1$ (bottom) formant spaces (in Hz) for the vowels and consonants pooled.

out by the controller (Abry *et al.*, 1994; Perrier *et al.*, 1996, 2000). In other words, any correlation observed between the articulatory variables should be used to reduce the number of degrees of freedom of the articulators. However, this approach must be carefully balanced by another criterion, the *biomechanical likelihood*. For instance, if larynx height and lip protrusion are inversely correlated due to the subject's articulatory control strategy (cf., e.g., Hoole and Kroos, 1998), two separate degrees of freedom should nevertheless be considered, even at the price of some residual correlation between the corresponding parameters.

2. Linear component analysis

Another important assumption in the present work is the *linearity* of the analysis and of the associated model: the shape data vectors *DT* are decomposed into linear combinations of a set of basic shape vectors *BV* weighted by loading factors *LF*, in addition to their average *neutral* shape $\overline{DT}$:

$$DT=\overline{DT}+LF\cdot BV.$$

Each loading factor *LFi* corresponds to an independent linear component, if its cross correlation with the other loadings is zero over the corpus of data. The dimensionality of the articulators' shapes and positions can thus be explored by classical linear analysis techniques such as principal component analysis (PCA) and linear regression analysis, as carried out by Maeda (1990, 1991), whose approach largely inspired the present work.

Maeda's approach to this decomposition was to iteratively determine each linear component in the following way: (1) the loading factor *LFi* is determined from the data as described below; (2) the associated basis shape vector *BVi* is determined by the linear regression of the current residual data for the whole corpus over *LFi*; (3) the corresponding contribution of the component is computed as the product of the loadings by the basis shape vector, and is finally subtracted from the current residue in order to provide the next residue for determining the next component.

For some of the linear components, the loading factors were arbitrarily chosen as the centered and normalized values of specific geometric measurements extracted from the contours, such as jaw or larynx height. For the other linear components, loading factors were derived by standard PCA applied to specific regions of the tongue contour.

Note that the solution of this type of linear decomposition is not unique in general: PCA delivers optimal components explaining the maximum data variance with a minimum number of components, but Maeda's linear component analysis allows a certain room of maneuver to control the nature and repartition of the variance explained by the components (for instance to make them more interpretable in terms of control), at the cost of a suboptimal variance explanation.

In this rest of this section, the various geometric measures are studied using statistical linear analysis in order to determine the correlations between these articulatory variables and to determine the degrees of freedom of the articulatory plant.

B. Jaw

The tongue is naturally identified as an important articulator in speech production, and its midsagittal contours, obtained from x-ray profile views of the vocal tract, have been the focus of most modeling efforts. The jaw has long been recognized as one of the main speech articulators, because it carries both the tongue and the lips. Its specific contribution to tongue shape has been clearly identified and related to the phonetic features of vowels (Lindblom and Sundberg, 1971). The dimensionality of jaw motion has been studied by many researchers (cf., e.g., Westbury, 1988; Edwards and Harris, 1990; Ostry *et al.*, 1997). The jaw, a rigid body, possesses six geometrical degrees of freedom (three rotations and three translations); however, it appears that for speech, movements are mostly restricted to the midsagittal plane if the rotation around the jaw axis is neglected (cf., e.g., Ostry *et al.*, 1997), which reduces the degrees of freedom of the jaw to three. From simple geometrical considerations, it is clear that the position of the jaw as a rigid body in a plane is uniquely defined by one rotation (defined here as *JawRot*) and by the two *x/y* translations of a reference point attached to the body, chosen as the upper edge of the lower incisors (defined here

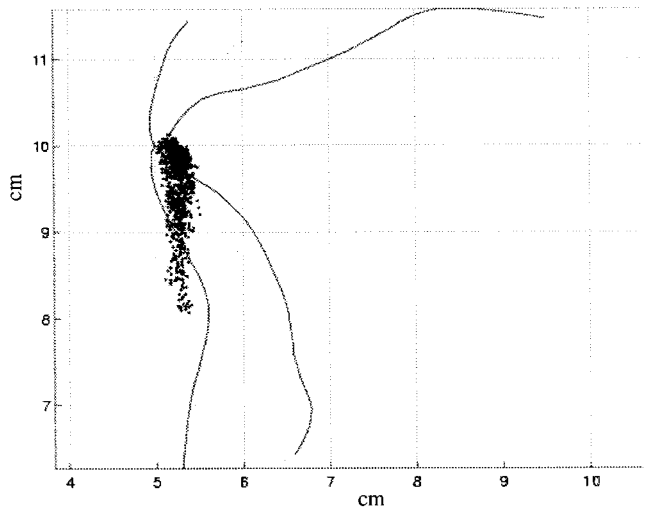

FIG. 3. Dispersion of the lower incisor upper edge superimposed on the contours of the hard palate and of the jaw.

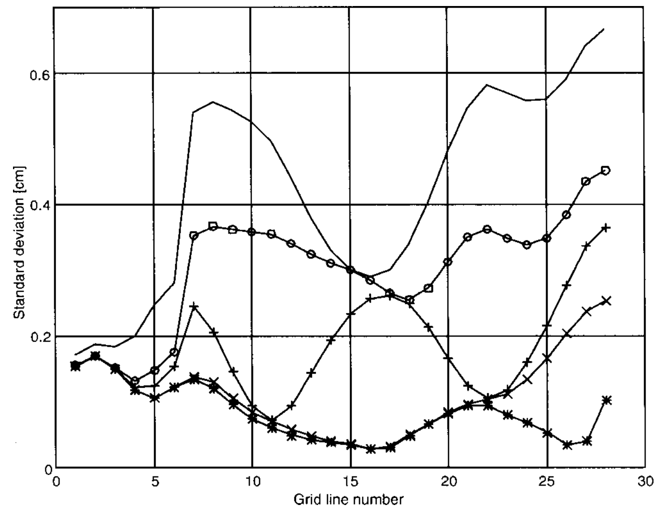

FIG. 4. Standard deviation (in cm) against gridline number for the inner contour of the vocal tract for the successive residues when the effect of the parameters are removed one by one (Raw data: solid line; data after suppression of the contribution of *JH* ''○,'' then of *TB* ''+,'' then of *TD* ''×,'' and finally of *TT* ''*'').

as *JawHei* and *JawAdv*; see Fig. 1). Figure 3 displays the resulting distribution of lower incisor position for the whole corpus.

PCA was applied to the centered—but nonnormalized—jaw position data *JawAdv* and *JawHei*. The first component explains 97.0% of the *JawAdv* and *JawHei* data variance. The overwhelming importance of this component could be predicted intuitively from the fact that the standard deviations of *JawAdv* and *JawHei* are, respectively, about 0.075 and 0.419 cm.

The jaw height component *JH* corresponding to the first degree of freedom of the jaw data was thus defined as the *JawHei* variable centered on its mean and normalized by its standard deviation. A second component, corresponding to jaw advance, *JA* was defined as the residue of *JawAdv* centered and normalized once the linear contribution of *JH* was removed. It can be concluded that for the present subject and corpus, the jaw possesses two independent degrees of freedom in the midsagittal plane, although the second component would have a rather limited influence on the tongue and the lips, as will be discussed further. The jaw was observed to be most retracted for labio-dentals: indeed this retraction allows the lower incisors and the upper lip to get in contact. Maximum jaw protrusion was observed for the coronal fricative [z]: this facilitates the creation of a constriction between the anterior region of the tongue blade and the front region of the alveolar ridge.

C. Tongue

The tongue shape is defined by the vector *Int* of the abscissa of its intersections with the grid lines. Since the jaw carries the tongue, the contribution of its movements should first be subtracted from tongue movements to maintain some biomechanical likelihood. However, due to the complexity of the muscular links between tongue and jaw (cf., e.g., Sanguineti *et al.*, 1998), it is very difficult to separate tongue movements induced by jaw movements from those due to active actions of tongue muscles themselves. In a study involving three subjects (including the subject of the present study), Bailly *et al.* (1998) showed that the slope of the regression line that links jaw height and tongue abscissa can be substantially greater than unity (by almost 100%). This means that tongue movements are apparently larger than the associated jaw movements, indicating that the subjects tend to actively move both jaw and tongue in synergy. In such a case, the passive tongue movement due to jaw movement needs to be determined. However, for the present subject, this synergy was rather weak compared to that of the other subjects (regression slopes lower than 1.15), and did not need to be taken into account: the *JH* parameter was directly considered as the first linear loading factor for each element of *Int*, and the corresponding prediction coefficients were obtained as the coefficients of the linear regression between *Int* and *JH* computed over all the items. Finally, the residual vector *Int_JH*, computed as the difference between predicted and measured values, for all the items, represents the tongue shape from which the contribution of the jaw has been removed.

The variance of the original *Int* data and the variance of the *Int_JH* residual data can be examined in Fig. 4 in terms of standard deviation (i.e., as the square root of the variance), as a function of grid line number. Table II gives, in addition, the global percentage of the total *Int* data variance explained by *JH* (numerical column 1).

The influence of the second jaw parameter *JA* upon tongue contours will be addressed later in this section.

The next step of the analysis consisted of extracting the degrees of freedom of the residual vector *Int_JH*. Gabioud (1994) showed that PCA applied to the whole tongue contour led to poor modeling of the tongue tip, even using three components. It was thus decided to apply PCA separately to the tongue body (gridline 7 to 24) and to the tongue tip (lines 24 to 28).

A first PCA procedure was thus applied to the residues of the 18 points considered for the tongue body, *Int_JH*(7:24). The first two components were retained. The

TABLE II. Summary of data variance explanation for the tongue contours. Column *Design* indicates how the factor was extracted. First column *Var* shows the ratio of data variance explained by the factor for the case the influence of jaw movements is taken into account by one parameter only. The second and third columns *Var* show the ratio of data variance explained when jaw is taken into account by two factors, the second factor being imposed at two different stages of the analysis.

Param.	*Design*	*Var.*	*Var.*	*Var.*
JH	Jaw height	52.2%	52.2%	52.2%
JA	Jaw advance		1.4%	
TB	PCA/tongue body	28.8%	28.6%	28.8%
TD	PCA/tongue body	11.4%	10.4%	11.4%
TT	PCA/tongue tip	3.6%	3.6%	3.6%
JA	Jaw advance			0.2%
	Total	**96.0%**	**96.1%**	**96.1%**

corresponding principal axes are characterized by the eigenvectors associated with the highest two eigenvalues of the cross-correlation matrix computed from these residues. The projections of the centered and normalized residues on these two principal axes give the values of the two associated components: *tongue body* component *TB*, and *tongue dorsum* component *TD*, which describe, respectively, the *front–back* and *flattening–arching* movements of the tongue (see also the nomograms in Fig. 8). These components were then used as predictors for the whole tongue contour. Table II presents a summary of the proportion of the total tongue data variance explained by each component, while Fig. 4 shows the details of the variance of the residues.

The tongue tip was found to possess two independent degrees of freedom [its coordinates, measured as *TngAdv* and *TngTip*[1] (see Fig. 1) are plotted in Fig. 5]: indeed, the residues of *TngAdv* and *TngTip*, after subtraction of the contributions of *JH*, *TB*, and *TD* (determined by the linear regression of *TngAdv* and *TngTip* for the whole corpus over *JH*, *TB*, and *TD*) are clearly not correlated. A first component, more generally dedicated to the representation of the apical region of the tongue, was then extracted: the *tongue tip* component *TT* is defined as the first component determined by the PCA of the residues of the tongue tip region (lines 24 to 28), from which the contributions of *JH*, *TB*, and *TD* have been removed. Its effects can be observed in Fig. 4 and Table II.

The *tonge advance* parameter *TA* was defined as the centered and normalized residue of the measured tongue advance *TngAdv* from which the contributions of *JH*, *TB*, *TD*, and *TT* were subtracted. Since it was, as expected, found to have a negligible predictive power on the tongue abscissa, it was not used as a loading factor for *Int*, but just to control the longitudinal extension of the grid in the front mouth region.

In order to test the influence of the *JA* parameter upon tongue contours, two experiments were carried out: in a procedure similar to that applied to *JH*, *JA* was used as the second imposed loading factor for the tongue analysis in one experiment, and as the loading factor imposed after *JH*, *TB*, *TD*, and *TT* in the other experiment. It was found that *JA* explained only 1.3% of the tongue data variance in the first

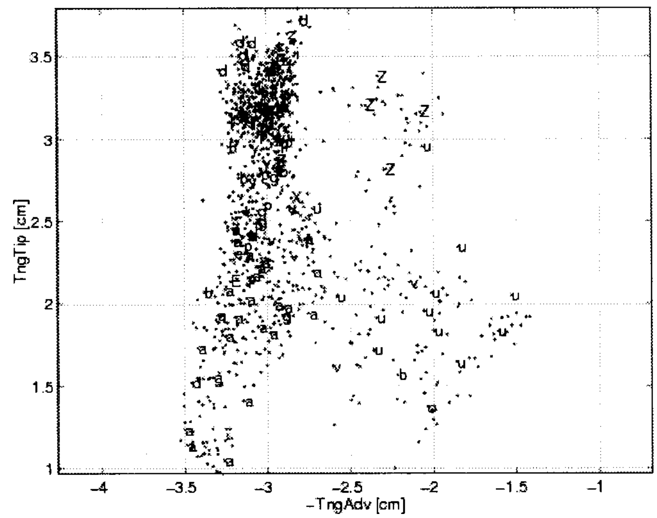

FIG. 5. Plot of the *TngAdv/TngTip* coordinates of the tongue tip (note that these are expressed in the rotated coordinate system attached to the gridline for the front part of the vocal tract; see Fig. 1). Phonemes /ø/, /ʒ/, /ɛ/ are referred to by symbols X, Z, and E, respectively.

case (see Table II, numerical *Var* column 2), and 0.2% in the second case (numerical *Var* column 3). A comparison of the associated nomograms in Fig. 6 suggests that the data variance explained by *JA* in the first case is actually explained by the other components *TB*, *TD*, and *TT*, in the second case. This hypothesis is also supported by results in Table II (and by a more detailed analysis of tongue shape data). *JA* was therefore not used as a control parameter of tongue shape.

In summary, the tongue contours in the grid line system possess four degrees of freedom, controlled by components *JH*, *TB*, *TD*, and *TT*. These four components account for 96% of the tongue variance data, which is only 1.5% less than the variance explained by the first four independent components (but with no direct articulatory interpretation) of a principal component analysis. The standard deviation of the residual error (normally distributed around zero on each gridline) reaches a maximum of 0.15 cm in the vicinity of the pharynx and of 0.1 cm at the tongue tip. The rms reconstruction error for the tongue, i.e., the root mean square error between the measured tongue data and the data calculated with the linear decomposition, amounts to a global value of 0.09 cm, while reaching maxima of 0.15 cm in the vicinity of the pharynx and of 0.1 cm at the tongue tip. The relatively poor modeling of the tongue tip extremity (which results in

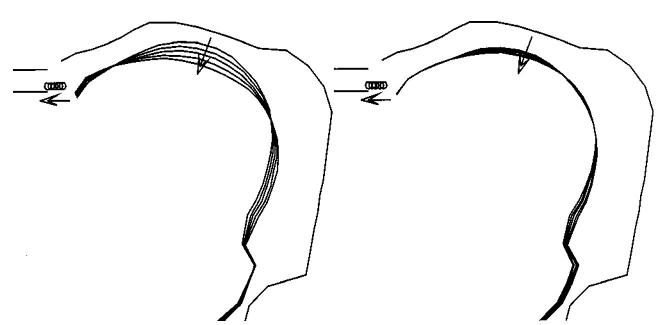

FIG. 6. Articulatory nomograms for *JA*. Left: case where *JA* is the second loading factor in the analysis; right: case where *JA* is the fifth loading factor after *JH*, *TB*, *TD*, and *TT*.

TABLE III. Correlation coefficients of some articulatory measurements. Absolute values higher than 0.6 are in bold face.

	JawHei	*JawAdv*	*LipHei*	*LipTop*	*ProTop*	*ProBot*	*TngTip*	*TngAdv*	*TngFlo*	*TngBot*	*LarHei*	*XHyoid*	*YHyoid*
JawHei	1.000												
JawAdv	−0.226	1.000											
LipHei	**0.641**	−0.039	1.000										
LipTop	−0.448	0.127	0.176	1.000									
ProTop	−0.463	0.148	−0.443	0.060	1.000								
ProBot	**−0.608**	0.062	−0.673	0.024	**0.912**	1.000							
TngTip	**−0.734**	0.207	−0.325	0.375	0.027	0.150	1.000						
TngAdv	0.412	−0.108	0.366	−0.242	−0.568	−0.571	0.142	1.000					
TngFlo	**−0.593**	0.053	−0.290	0.219	−0.083	0.056	**0.877**	0.454	1.000				
TngBot	−0.038	−0.109	−0.347	−0.325	0.352	0.411	−0.219	−0.205	−0.160	1.000			
LarHei	−0.358	−0.068	**−0.550**	−0.136	**0.574**	**0.664**	0.005	−0.380	−0.008	0.797	1.000		
XHyoid	**0.733**	−0.128	0.544	−0.259	−0.416	−0.533	−0.556	0.288	−0.469	−0.169	−0.421	1.000	
YHyoid	0.355	0.100	0.522	0.099	−0.460	−0.572	−0.047	0.351	−0.033	−0.815	**−0.859**	0.574	1.000

only minor acoustical effect) is mainly due to measurement inaccuracies related to the difficulty of precisely defining this tongue tip extremity. A supplementary articulatory control parameter could be extracted to more precisely control the pharyngeal region as implied in the ±Advanced Tongue Root languages (cf., e.g., Tiede, 1996).

Recall finally that the grid system is controlled, in addition, by two parameters, i.e., *TA*, and a parameter related to *LarHei* that will be defined in Sec. III E.

D. Lips

A PCA analysis revealed that 98.4% of the variance of the lip measures *LipHei*, *LipTop*, *ProTop*, and *ProBot* can be explained by three independent components, in addition to the natural contribution of jaw height to lip shape. This is expected, as lip protrusions *ProBot* and *ProTop* are strongly correlated (cf. Table III). Note also that in another study on the same subject, where the lip shape was more accurately described as a three-dimensional mesh of points controlled by the 3-D coordinates of 30 control points (Revéret and Benoît, 1998), Badin *et al.* (2000) also found that three degrees of freedom were sufficient to describe the position of the lips on a corpus of 34 sustained articulations (French vowels and consonants), in addition to the *JH* contribution (the *JA* contribution explained only 1% of the lip data variance). These degrees of freedom are related to three gestures: lip protrusion/rounding, lip closure, and a sort of simultaneous vertical movement of both lips as needed for the subject to realize labio-dentals. In order to simplify the model and its relations to simple articulatory measurements and acoustic interpretations of the lip horn, we decided to use an equivalent set of components: (1) a component related to *LipHei*, taken into account by *LH*, the centered and normalized residue of *LipHei* after removing the *JH* contribution; (2) a component related to *ProTop*, *LP*, the centered and normalized value of the residue of *ProTop* after removing the *JH* contribution; and (3) a component related to a mere vertical, roughly synchronous, movement of both upper and lower lips relative to upper incisors lower edge, taken into account by the *lip vertical position* parameter *LV*, the centered and normalized residue of *LipTop* after removing *JH*, *LH*, and *LP* contributions. Note that this approach results in a slight correlation between *LP* and *LV*. Note also that the horizontal jaw retraction aiming at producing labio-dental constrictions is not taken into account as such, but that its acoustical consequences are dealt with in an indirect way (cf. Sec. IV B 3).

E. Other articulatory measurements

Finally, a number of other articulatory measurements were analyzed. Table III provides the linear correlation coefficients between these measurements.

Table III shows that *LarHei* is partially correlated with *ProBot*, *ProTop*, and *LipHei*. These correlations cannot be explained *a priori* by obvious biomechanical effects, and will thus be ascribed to the speaker control strategies. Indeed, it is clearly established that lip rounding and larynx lowering constitutes, for some subjects, a synergetic strategy for high rounded vowels [uy] (Hoole and Kroos, 1998). Larynx height was thus represented by its centered and normalized value *LY*, and further used to control the grid system (cf. Sec. IV A 2).

The horizontal position of the hyoid bone, *XHyoid*, is very highly correlated to jaw height, while its vertical position, *YHyoid*, is even more strongly correlated to larynx height (see Table III and Fig. 7). These two components are

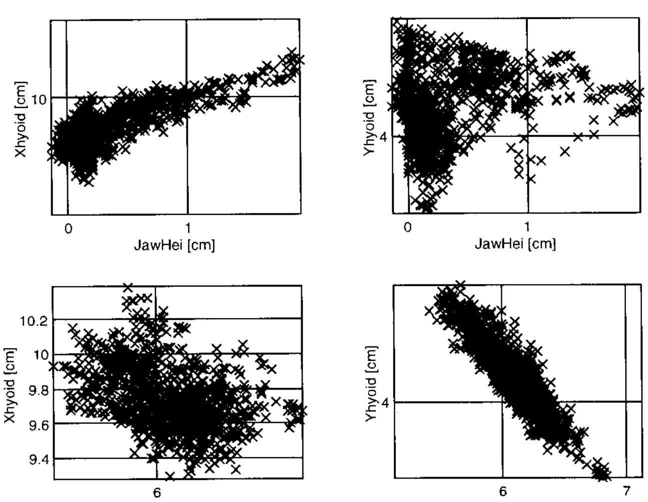

FIG. 7. Plots showing the relations between hyoid bone coordinates and jaw and larynx heights.

TABLE IV. Correlation coefficients of the articulatory control parameters of the model.

	JH	*TB*	*TD*	*TT*	*TA*	*LY*	*LP*	*LH*	*LV*
JH	1.000								
TB		1.000							
TD			1.000						
TT				1.000					
TA					1.000				
LY	−0.358	−0.041	0.272	−0.499	−0.103	1.000			
LP		0.215	0.472	−0.125	−0.118	0.461	1.000		
LH		−0.256	−0.039	0.236	−0.075	−0.417	−0.215	1.000	
LV		0.215	0.065	−0.006	−0.164	−0.045			1.000

clearly less correlated with each other (correlation coefficient $R=0.574$), than found by Westbury (1988) using a more restricted corpus for a single subject ($R=0.871$).

Note that the position of the highest connection point between tongue and epiglottis, referred to as *TngBot* (see Fig. 1), is highly correlated with *LarHei*, as expected. The elevation of tongue floor *TngFlo* is correlated with *TngTip* and *JawHei*.

IV. BERGAME: AN ARTICULATORY-ACOUSTIC MODEL

As stated above, the main function of an articulatory model is to offer a compact representation of articulation, i.e., a representation that needs as few control parameters as possible and is nevertheless accurate enough to be meaningful for speech. The analysis presented in the previous section prepared the ground for establishing such a model, which is necessarily the result of a compromise between a minimum number of control parameters and a maximal explanation of the data variance (or minimal data reconstruction error). The present section describes *Bergame*, an articulatory-acoustic model developed at ICP with the aim of mimicking as closely as possible experimental data gathered on the reference subject.

Bergame consists of: (1) a physiologically oriented linear articulatory model, based on the articulatory data measured from the cineradiofilm and the video labiofilm made on the reference subject; (2) a model of midsagittal-to-area function conversion based on the same subject; (3) an acoustic model.

A. The linear articulatory model

The principle of a linear articulatory model is to *calculate the position and shape of the various articulators as linear combinations of the articulatory control parameters.* The development of the model thus amounts to defining the control parameters and to determining the coefficients of these linear combinations. The nine parameters chosen for controlling the articulatory model stem directly from the previous component analysis: *JH*, *TB*, *TD*, *TT*, *TA*, *LY*, *LH*, *LP*, and *LV*, which are dimensionless, centered, and normalized. These parameters are, in most cases, orthogonal to each other, the exceptions (see Table IV) being due to the subject and language specific control strategies. The model equations are described in some detail in the following. The model behavior is illustrated in Fig. 8 by *articulatory nomograms*, i.e., the variations of the midsagittal contours resulting from variations of the articulatory control parameters from -3 to $+3$ with $+1$ steps.

1. Jaw

The jaw has been shown above to possess essentially one degree of freedom for this subject and the corpus analyzed. Jaw position is therefore controlled by the single parameter *JH* that defines *JawHei_mod* by the simple linear relation:

$$JawHei\_mod = JawHei\_mean + JawHei\_std \cdot JH,$$

where *JawHei_std* is the standard deviation of *JawHei* and *JawHei_mean* its mean over the corpus.

2. Tongue, midsagittal distances, and vocal tract outer contours

Since the tongue contours are attached to the grid lines, the next necessary step is to determine the position of the mobile parts of the grid system, namely *TngAdv* and *LarHei*. The modeled tongue advance, *TngAdv_mod*, was found to be almost linearly related to *TA*, *JH*, *TB*, and *TD*, and was therefore controlled by:

$$TngAdv\_mod = TngAdv\_mean + pred\_TngAdv\_JH\_TB\_TD\_TA \cdot [JH, TB, TD, TA],$$

where [*JH,TB,TD,TA*] is the matrix of control parameters, and *pred_TngAdv_JH_TB_TD_TA* are the associated coefficients determined by multiple linear regression. As seen in Sec. II E, *LarHei* is controlled only by *LY*, and not *LP* and *LH*, despite a slight correlation between lips and larynx, in order to ensure an independent control of lips and larynx in the model. *LY* is therefore partially correlated with a number of other control parameters, as seen in Table IV. Note that *TngAdv* and *LarHei* are reconstructed without error.

Finally, the abscissa of the whole tongue contour *Int_mod* (lines 1 to 28) is determined as linear combinations of the parameters *JH*, *TB*, *TD*, and *TT*:

$$Int\_mod = Int\_mean + pred\_Int\_JH\_TB\_TD\_TT \cdot [JH, TB, TD, TT].$$

Figure 8 displays articulatory nomograms for *JH*, *TB*, *TD*, *TT*, and *TA* as well.

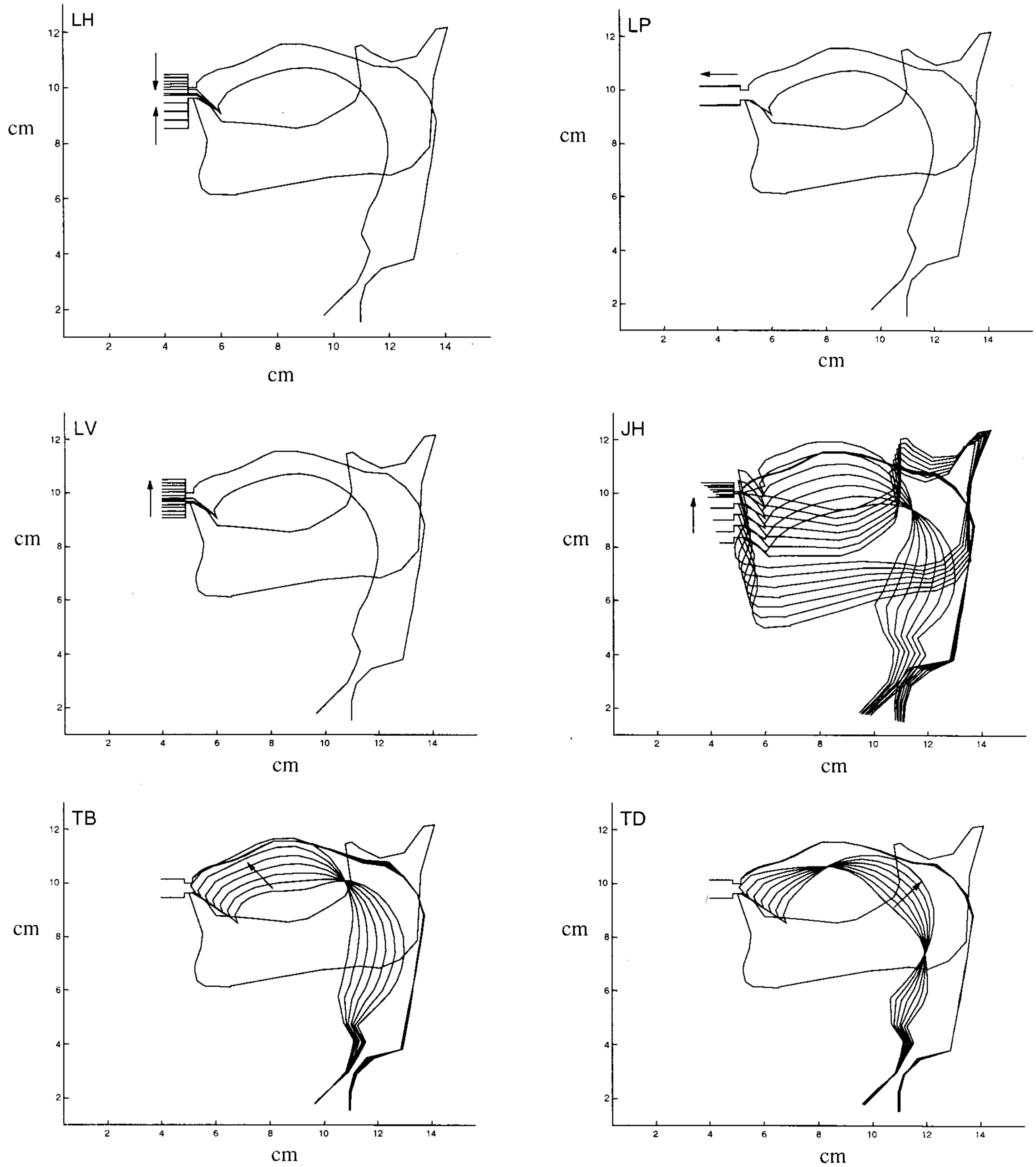

FIG. 8. Articulatory nomograms: variations of the midsagittal contours resulting from variations of the articulatory control parameters from -3 to $+3$ with $+1$ steps. Note that the movements of upper and lower lips are *opposite* for the *LH* nomogram, but *parallel* for the *LV* nomogram.

The midsagittal distances have been similarly handled. The abscissa of the vocal tract outer contours are computed as the sum of the abscissa of the tongue and the corresponding midsagittal distances, except for the hard palate region that is considered as a fixed contour.

3. Lips

It has been shown above that the lip geometry of the subject is best described with three degrees of freedom, represented by *LH*, *LP*, and *LV*, in addition to the contribution of *JH*. The lip horn, considered as the vocal tract region anterior to the upper incisor plane, is represented by a single tube section, with a length *ProLip_mod* proportional to the prediction of the *LipTop* dimension, with a proportionality factor of 0.6. This length reduction aims at approximately taking into account the fact that the lip corner position is not known, and that the effective acoustical end of the lip horn is located between the lip corner and the extremities of the lips measured by *ProTop* and *ProBot*. The parameter *ProLip_mod* is thus defined by:

$$ProLip\_mod = 0.6 \cdot (ProTop\_mean + pred\_ProTop\_JH\_LH \cdot [JH, LH]),$$

where the coefficients *pred_ProTop_JH_LH* are obtained by multiple linear regression.

Lip height *LipHei_mod* is similarly modeled as a linear combination of *JH*, *LH*, and *LP*, as well as lip width *A_mod*. Lip vertical position *LipTop_Mod* is also a linear combina-

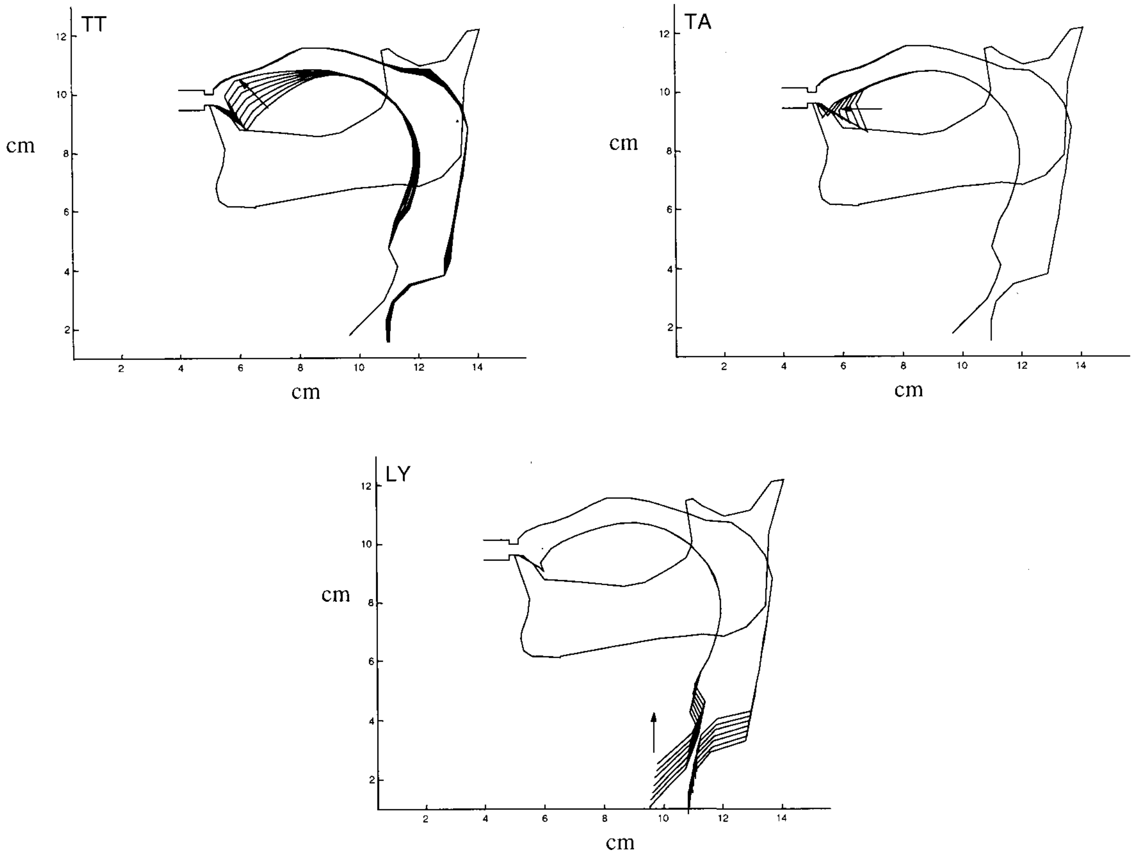

FIG. 8. (*Continued.*)

tion of *JH*, *LH*, *LP*, and *LV*. Figure 8 shows also the articulatory nomograms relating these lip parameters.

B. Midsagittal and area functions—Acoustic models

The midsagittal contour alone is not sufficient to derive the corresponding vocal tract acoustic features. Acoustic vocal tract models are indeed based on tube acoustics, and thus need a description of the vocal tract in terms of *area function*. Most studies devoted to the problem of converting midsagittal distances d to area functions S resulted in solutions based on the ''α,β model'' proposed by Heinz and Stevens (1965), where $S=\alpha\cdot d^{\beta}$: the principle consists in calculating the area of each vocal tract section as a power function of the corresponding midsagittal distance (cf., e.g., Beautemps *et al.*, 1995, for more details). Finally, vocal tract aero-acoustic simulations in the time or in the frequency domain allow the computation of the speech signal or speech acoustic characteristics from the area function (cf., e.g., Maeda, 1982; Badin and Fant, 1984; Mawass *et al.*, 2000).

1. Vocal tract

The midsagittal function represents the sagittal distances between the tongue contour and the outer vocal tract contour along the vocal tract midline, estimated for each section enclosed between two consecutive measurement grid lines. For each section, a quadrilateral can be defined in the midsagittal plane by the intersection points of the tongue contour and of the vocal tract outer contours with the corresponding two lines of the grid. The midsagittal distance for this section is calculated as the surface of the quadrilateral divided by the length of the section. Following Heinz and Stevens (1965), the vocal tract area function is estimated from the midsagittal function. It uses an extended version of a conversion model (Beautemps *et al.*, 1996) optimized for both vowels and consonants: the area function is then derived from the midsagittal function using a polynomial expression where the cross-sectional area S depends on both the midsagittal distance d and the x distance from the glottis measured along the vocal tract midline:

$$S(x,d)=\alpha_1(x)\cdot d+\alpha_2(x)\cdot d^{1.5}+\alpha_3(x)\cdot d^2+\alpha_4(x)\cdot d^{2.5}.$$

The $\alpha_i(x)$ functions are expressed as Fourier series, up to the third order, of $\pi.x/l_{\mathrm{tot}}$, where the l_{tot} is the vocal tract length (including the lips):

$$\alpha_i(x)=a_{i0}(x)+\sum_{n=1}^{3} a_{in}\cdot\cos\left(n\frac{\pi}{l_{\mathrm{tot}}}x\right)+\sum_{n=1}^{3} b_{in}\cdot\sin\left(n\frac{\pi}{l_{\mathrm{tot}}}x\right).$$

The values of the Fourier coefficients (altogether, 28 parameters) were optimized so as to minimize, for the N selected configurations, the χ^2 distance between the four formants F_{ik} computed from the area function derived from the synthesized contours and the formants F_{ik}^c measured on the acoustic signal of the original data:

$$\chi^2=\sum_{i=1}^{N}\sum_{k=1}^{4}\frac{(F_{ik}^c-F_{ik})^2}{F_{ik}^c}.$$

At first, the optimization procedure was applied to a restricted set of eight vowels [aɛeiyuoø], in order to ensure an easy convergence. The results were then refined by applying the same optimization procedure to the whole corpus, excluding only the data for which measurement of the four formants was not possible, i.e., excluding the 347 configurations mainly associated with occlusive consonants [pbdg].

Dang and Honda (1998) developed a similar polynomial decomposition where the coefficients, a function of the distance from the glottis, are determined by minimizing the difference between the estimated and MRI-based area functions for five Japanese vowels. The present procedure, developed before any 3-D vocal tract data were available for the subject, does not make use of 3-D data. However, Badin *et al.* (1998) subsequently acquired 3-D MRI data allowing the direct determination of both midsagittal contours and area functions for a set of vowel articulations for the same subject. These data have therefore been used to assess the quality of this algorithm: for the ten French vowels [aɛeiyuoɔøœ] the comparison between the areas directly estimated from the 3-D measurements (excluding the larynx region and the lips for which no MRI data were available) and those computed by the present procedure from the midsagittal contours estimated from the 3-D measurements has revealed a global root mean square (rms) error value lower than 0.6 cm^2. The relatively important rms error (more than 1 cm^2) observed in the low pharyngeal region for [aɔøœ] is probably due to the whispered production mode used to maintain the articulation during the long 3-D MRI data recording duration whose main consequence is a more constricted tongue in the back region (cf. Matsuda and Kasuya, 1999). This implies a decrease of the cross-sectional areas between the glottis and the epiglottis and probably modifies the relation between the midsagittal distances and the related area functions. Finally, in the uvular region, the small midsagittal distance measured is not representative of the entire cross-section, due to the fact that the main part of the velum body is concentrated in the midsagittal plane with free air flow on both sides. The consequence is an underestimation of the area inherent to the conversion model.

The fit between measured and reconstructed data was also assessed at the level of midsagittal functions. The midsagittal functions of the synthesized vocal tract contours have thus been compared to the midsagittal functions of the original data calculated with the measurement gridline system implemented in the model. The rms errors of the length and of the midsagittal distances are almost zero except for the larynx region where the errors can respectively reach 0.1 and 0.3 cm, probably due to the poor modeling of the tongue in this region (see Fig. 4). A maximum of 0.2 cm for the rms error on the midsagittal distances is also noted in the front part of the vocal tract between the last tongue point and the teeth, due to the fact that the sublingual cavity is not taken into account in the model.

2. Lip area

A possibility for computing lip area *S_mod* is the relation established for the first time by Fromkin (1964):

$$S\_mod = pred\_S\_A\_B \cdot A \cdot B,$$

where *A* and *B* are, respectively, the intra-oral lip width and height measured from the video labio-film. For the present

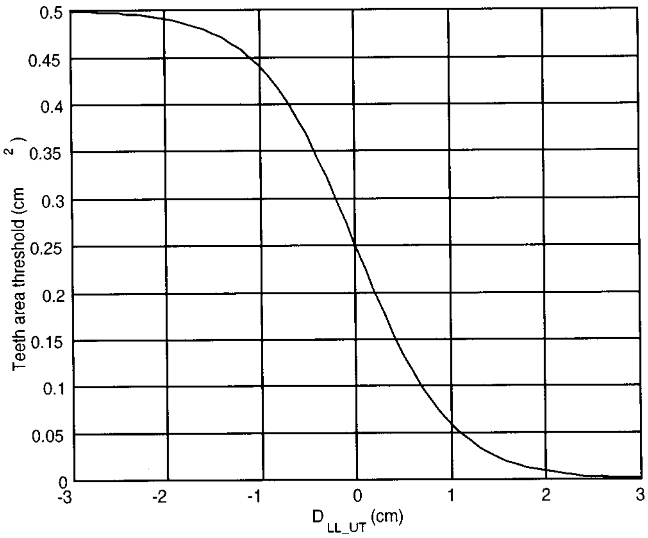

FIG. 9. Minimum threshold for the area at the incisors as a function of difference between lower lip position and upper incisor edge.

subject, a value of 0.80 was found for *pred_S_A_B* by linear regression applied to the whole corpus. However, the lack of accuracy of *A_mod* resulted in a poor modeling of small areas. This method was therefore abandoned, and lip area was in practice calculated as a second order multilinear regression of the *JH*, *LH*, and *LP* components, for which the coefficients were optimized as to obtain the best fit to the lip area measured on video front pictures. With this modeling, we obtained 0.2 cm^2 for the rms error.

3. Acoustic effect of LV

In the absence of the horizontal jaw control parameter *JA*, only vertical movements are taken into account for the lower incisors, i.e., through *JH*. Therefore, there is no straightforward provision in the model for producing labio-dental constrictions by a combination of jaw retraction and lower lip elevation movements, which is the standard articulatory strategy for producing labio-dental fricatives.

This problem is overcome by a mechanism that uses the *LV* parameter for the production of the labio-dental constriction at the incisor section. The incisor section area is made an indirect function of lower lip vertical position, and thus of *LV*, by limiting it to a minimum threshold value function of the difference between lower lip position and upper incisor edge (see Fig. 9). This allows *LV* to be audible, i.e., to have acoustic consequences, at least in circumstances typical of labio-dentals where the lower lip has to be higher than the upper incisor edge in order to produce the proper constriction. This feature was particularly useful for the inversion of the articulatory-to-acoustic relation for fricatives (Mawass *et al.*, 2000).

4. Acoustic model

Acoustic transfer functions as well as formants and bandwidths were determined from these area functions by means of a frequency domain vocal tract acoustic model (Badin and Fant, 1984). A time domain reflection-type line analogue (Bailly *et al.*, 1994), extended to include improved

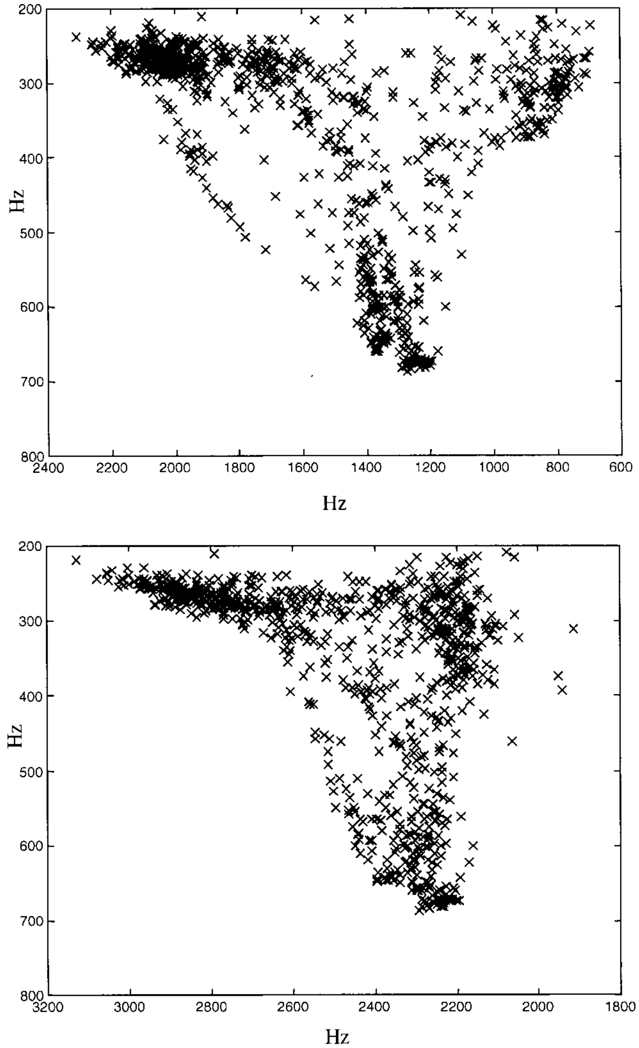

FIG. 10. Predicted $F2/F1$ (top) and $F3/F1$ (bottom) formant spaces (in Hz) for the vowels and consonants pooled.

voice (Pelorson *et al.*, 1996) and noise source models (Badin *et al.*, 1995), can also be driven by these area functions, in association with lung pressure and vocal cords parameters, to produce high quality articulatory synthesis (cf Mawass *et al.*, 2000).

The area functions and the derived formants have been computed for the assessment of the midsagittal to area function conversion, starting from the synthesized midsagittal contours. The rms error and the rms relative error on formants have both been calculated for the whole corpus (excluding 347 configurations for which the measurement of the four formants was not possible): 45 Hz (12.86%), 100 Hz (7.32%), 162 Hz, (6.47%), and 173 Hz (5.21%), respectively, for $F1$, $F2$, $F3$, and $F4$. The mean differences between the formants obtained by the model and those measured are 14 Hz, -65 Hz, and 26 Hz, respectively, for $F2$, $F3$, $F4$ (no significant difference was found for $F1$, except 27 Hz for the vowels).

The predicted maximal formant spaces are comparable to the measured ones (cf. Fig. 2 and Fig. 10). However, the computed formant $F1$ of [a] is about 34 Hz too low.

When modeling the area function of the four point vowels of their two American subjects, Baer *et al.* (1991) reported a deviation (in terms of the rms of the relative error) of 13%, 31%, and 13% on the measures for, respectively, $F1$, $F2$, and $F3$; these deviations are noticeably higher than those in the present study. From a set of 16 disyllabic utterances [hə'CV] and one sentence, Mermelstein (1973) obtained 10.3%, 4.9%, and 5.5% for the average absolute error on $F1$, $F2$, and $F3$; in terms of rms, these errors are still lower. Mermelstein's fits are thus clearly better than ours; however, it should be recalled that they were obtained on a much more restricted set of data.

V. DISCUSSION AND PERSPECTIVES

A. Summary

A linear component analysis of tongue contours and articulatory measures extracted from cineradio- and labio-films made on a reference subject revealed a relatively small number of degrees of freedom. The jaw appears, for the subject studied, to have mainly two degrees of freedom, related to the lower incisors vertical and horizontal movements. However, only the vertical component exerts a significant effect on tongue shape. The residue of tongue shape, once the contribution of the jaw has been removed, possesses four degrees of freedom: tongue body, tongue dorsum, tongue tip, and tongue advance. An extra parameter takes into account the larynx height variance. Similarly, the lip shape possesses, in addition to the jaw contribution, three degrees of freedom: lip protrusion, lip height, and lip vertical elevation. These nine parameters are mostly independent of each other (cf. Table IV), except for *LY* that is correlated with lips and tongue parameters and the correlation between *LP* and *TD*. These degrees of freedom are specific to the vocal tract and articulators of one subject uttering one specific corpus in one language. The corpus was designed to include as many French vowels and consonants as possible. A linear articulatory model was developed based on these data; it explains 96% of the tongue data variance, with an rms reconstruction error of about 0.09 cm. It was complemented by a model converting the midsagittal contours to an area function based on a fitting of midsagittal functions and formant frequencies. Finally this model allows the calculation of formants with rms errors of 45 Hz for $F1$, 100 Hz for $F2$, and 162 Hz for $F3$ over the corpus. To the knowledge of the authors, no such comprehensive model has been developed so far; most of the available models deal with vowels only, while others do not include acoustics.

B. Choice of subject and corpus

The development of such an articulatory-acoustic model based on a specific reference subject was motivated by the need for a model that could fit a real subject's midsagittal profiles of French fricative consonants, plosives, and vowels, as well as formants, with a fairly high degree of accuracy for a large number of configurations. The possibility now exists to investigate in detail the articulatory strategies employed by the subject, and in particular coarticulatory strategies (cf. Mawass *et al.*, 2000, or Vilain *et al.*, 1998). One may argue that no general conclusions may be drawn from such studies, as they are supported by one single subject's data. However, we were very much aware of the risk of blurring out clear individual articulatory strategies employed by individual subjects when merging together several subjects' data, and therefore made the choice of a single subject for the present study. Similar analyses are under way for other subjects, in order to determine which features may be considered as general and which ones as more subject-specific (Bailly *et al.*, 1998; Vilain *et al.*, 1998; Engwall and Badin, 1999). These studies will also allow us to investigate the number of degrees of freedom of the jaw involved in speech.

The influence of the number of items used for the linear analysis was studied for the present subject by Badin *et al.* (1998); they found that, by choosing the contour samples, i.e., by selecting only vowel and consonant targets in the initial corpus, an articulatory model was produced that represented the whole corpus data with an accuracy close to that obtained when the full model based on the whole corpus was used. More specifically, they showed that the data reconstruction error, computed as the rms error of the abscissa of the tongue contour along each grid line for the 1222 images of the available corpus of midsagittal contours, was 0.09 cm, 0.11 cm, and 0.17 cm when the model was elaborated using, respectively, 1222, 20, and 8 configurations. This justifies the elaboration of models from a much lower number of articulations, and thus, in particular, the use of MRI images instead of x-ray images.

C. Comparison of the degrees of freedom found in other studies

Degrees of freedom are clearly subject and corpus-dependent but their number and their definition are closely related to the method used to explain the whole data set variance. The linear component analysis used by Maeda (1979, 1990) is sometimes referred to as a two-way factor analysis of the variance, where one mode corresponds to the predictors and the other one to the matrix of coefficients of the linear combinations. Using this principle, Maeda (1979) extracted one loading factor for the jaw, and three for the residual midsagittal tongue data to explain 98% of the variance for a corpus made of 400 frames of [pV_1CV_2] ([aiu] and [dg]) sequences uttered by one subject. In an extended corpus of 519 frames corresponding to 10 French sentences, three supplementary components were obtained for the lips including the frontal lip-opening shapes, and four tongue degrees of freedom explaining 88% of the variance (Maeda, 1990). Finally, for these data, Sanguineti *et al.* (1998) imposed two degrees of freedom for the jaw (protrusion and rotation), and one for the larynx, and obtained three other degrees of freedom for the tongue residue from a similar analysis in the so-called λ-space.

The PARAFAC method (Harshman *et al.*, 1977; Nix *et al.*, 1996; Hoole, 1999) is a three-way factor generalization where the third mode corresponds to linear coefficients that account for differences between subjects. Harshman *et al.* (1977) derived two components for the description of representative midsagittal tongue and lip shapes of ten English vowels uttered by five subjects. Hoole (1999) proposed a two-factor PARAFAC model of the German vowels in a symmetrical stop consonant context, plus an additional PCA component to capture the subject-specific nonvocalic behavior of the tongue. More recently, Hoole *et al.* (2000) extracted a two-factor PARAFAC solution that explained 90% of the variance from a set of MRI midsagittal tongue contours measured during the production of seven German vowels by nine speakers. The first component captured the dimension low-back to high-front, and the second was associated with the mid-front to high-back motion. The complex effect of the first component can be decomposed in a co-variation of *JH* and *TB*, the second component being related to *TD*. Ultimately, and to the knowledge of the authors, no analysis based on the PARAFAC method has been realized with an imposed jaw component. To conclude, no analysis based on PARAFAC principles has shown success in explaining the large and phonetically varied data obtained from multiple speakers.

Bailly *et al.* (1998) studied the synergy between tongue and jaw for three subjects, including the present reference subject. They found that the two other subjects used a fairly strong synergy: the amplitude of the tongue movements measured at the tip and at the root that were correlated with jaw movements were about twice as large as might be expected from the simple mechanical carrying effect of the jaw. In other words, the jaw and the tongue shared the execution of the tongue movements. The present subject does not use this synergy: the tongue is not so active, and appears to be passively carried by the jaw. However, all three subjects' articulators had qualitatively the same degrees of freedom. This synergy is still a crucial issue for understanding coarticulation strategies.

D. Perspectives

The principles of this work have been duplicated for the modeling of Swedish midsagittal tongue shapes (Engwall and Badin, 1999). Over 90% of the variance is explained by the four tongue degrees of freedom JH, TB, TD, and TT.

One of the main issues in the analysis of speech degrees of freedom is the possibility to build a linear articulatory model that takes into account the redundant feature of the articulators shapes. For instance, it can help to reconstruct complete tongue shapes from a reduced number of articulatory measurement points, such as those provided by electromagnetic articulometry. Badin *et al.* (1997) used the present model to retrieve, from one coil on the lower incisor and three coils on the tongue of the reference subject, the tongue shape as well as the articulatory control parameters *JH*, *TB*, *TD*, *TT*, and *TA* with a fairly good accuracy. This may be useful for investigating speech coarticulation and synergetic strategies (cf., e.g., Vilain *et al.*, 1998), for testing hypotheses of the Frame/Content concept in the child's language development (Vilain *et al.*, 1999), or evaluating the adaptability of speech articulation to various linguistic tasks and environmental conditions such as changes illustrated by the Lombard reflect (Beautemps *et al.*, 1999).

The present articulatory-acoustic model can also be used to derive, by inversion, articulatory control parameters from formants measured in other utterances produced by the same subject (Mawass *et al.*, 2000). These data, in conjunction with aerodynamic data obtained for the same subject, have been used for the articulatory synthesis of French fricatives (Mawass *et al.*, 2000).

Another extension of the present study is the third dimension. 3-D MRI images have been recorded for the same subject, and a 3-D linear articulatory model is being developed according to the same approach (Badin *et al.*, 1998, 2000); the new model has been elaborated in such a way that part of its control parameters are identical with those of the present midsagittal model, which opens the possibility of inheriting knowledge already acquired for the midsagittal plane, while acquiring new features such as the capability of producing lateral consonants. Finally, the modeling of the velum from MRI midsagittal data should complement the present model.

ACKNOWLEDGMENTS

This work has been partially funded by the European Community (ESPRIT/BR project *Speech Maps* No. 6975), and by the Rhône-Alpes Agency for Social and Human Sciences (ARASSH) (project ''A Virtual Talking Head: Data and models in speech production''). It has benefited from the valuable help of many people to whom the authors are very much indebted: Bernard Gabioud (who initiated a part of this work in the framework of the *Speech Maps* project), Tahar Lallouache (for the lip measurements), Shinji Maeda (for the initial version of the contour edition program, and more importantly for having largely inspired this work), Gilbert Brock, Péla Simon, and Jean-Pierre Zerling (for their expertise on cineradiography), Agnes Hennel (for access to the cineradiography equipment at the Strasbourg Schiltigheim Hospital), Thierry Guiard-Marigny and the late Christian Benoît (for their help on data gathering and processing), Christian Abry (for many stimulating discussions), Marija Tabain (for polishing our French English), as well as many other colleagues at ICP, Grenoble. We have also greatly appreciated the pertinent comments and careful editorial advice of Anders Löfqvist and two anonymous reviewers.

[1]Note that *TngTip* is identical to the last point of the tongue contour abscissa vector.

Abry, C., Badin, P., and Scully, C. (**1994**). ''Sound-to-gesture inversion in speech: The *Speech Maps* approach,'' in *Advanced Speech Applications*, edited by K. Varghese, S. Pfleger, and J. P. Lefèvre (Springer, Berlin), pp. 182–196.

Badin, P., Baricchi, E., and Vilain, A. (**1997**). ''Determining tongue articulation: from discrete fleshpoints to continuous shadow,'' in *Proceedings of the 5th EuroSpeech Conference* (University of Patras, Wire Communication Laboratory, Patras, Greece), Vol. 1, pp. 47–50.

Badin, P., Mawass, K., and Castelli, E. (**1995**). ''A model of frication noise source based on data from fricative consonants in vowel context,'' in *Proceedings of the 13th International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud (Arne Strömbergs Grafiska Press, Stockholm, Sweden), Vol. 2, pp. 202–205.

Badin, P., Bailly, G., Raybaudi, M., and Segebarth, C. (**1998**). ''A three-dimensional linear articulatory model based on MRI data,'' in *Proceedings of the Third ESCA/COCOSDA International Workshop on Speech Synthesis* (Jenolan Caves, Australia), pp. 249–254.

Badin P., Motoki, K., Miki, N., Ritterhaus, D., and Lallouache, T. M. (**1994a**). ''Some geometric and acoustic properties of the lip horn,'' J. Acoust. Soc. Jpn. (E) **15**, 243–253.

Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M., and Segebarth, C. (**2000**). ''Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images,'' 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. 261–264.

Badin, P., and Fant, G. (**1984**). ''Notes on vocal tract computation,'' Speech Transmission Laboratory—Quarterly Progress Status Report Vol. 2-3/1984, pp. 53–108.

Badin, P., Shadle, C. H., Pham Thi Ngoc, Y., Carter, J. N., Chiu, W., Scully, C., and Stromberg, K. (**1994b**). ''Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis,'' in *Proceedings of the 3rd International Conference on Spoken Language Processing* edited by Mike Edington (Yokohama, Japan), Vol. 1, pp. 163–166.

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**, 799–828.

Bailly, G. (**1993**). ''Resonances as possible representations of speech in the auditory-to-articulatory transform,'' in *Proceedings of the 3rd Eurospeech Conference on Speech Communication and Technology* (Berlin), Vol. 3, pp. 1511–1514.

Bailly, G., Badin, P., and Vilain, A. (**1998**). ''Synergy between jaw and lips/tongue movements: Consequences in articulatory modelling,'' in *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by R. H. Mannell, J. Robert-Ribes, and E. Vatikiotis-Bateson (Australian Speech Science and Technology Association, Inc., Sydney, Australia), Vol. 5, pp. 1859–1862.

Bailly, G., Castelli, E., and Gabioud, B. (**1994**). ''Building prototypes for articulatory speech synthesis,'' in *Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis* (New York), pp. 9–12.

Beautemps, D., Badin, P., and Laboissière, R. (**1995**). ''Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data,'' Speech Commun. **16**, 27–47.

Beautemps, D., Borel, P., and Manolios, S. (**1999**). ''Hyper-articulated speech: Auditory and visual intelligibility,'' in *Proceedings of the 6th European Conference on Speech Communication and Technology*, Vol. 1 (Budapest, Hungary), pp. 109–112, September 1999.

Beautemps, D., Badin, P., Bailly, G., Galván, A., and Laboissière, R. (**1996**). ''Evaluation of an articulatory-acoustic model based on a reference subject,'' in *Proceedings of the 4th Speech Production Seminar* (Autrans, France), pp. 45–48.

Boë, L. J., Gabioud, B., Schwartz, J. L., and Vallée, N. (**1995**). ''Towards the unification of vowel spaces,'' in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud (Arne Strombergs Grafiska Press, Stockholm, Sweden), Vol. 4, pp. 582–585.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P. (**1986**). ''Cinéradiographie des voyelles et consonnes du francais [Cineradiography of vowels and consonants in French],'' Trav. de l'Inst. de Phonétique de Strasbourg, 296 pp.

Coker, C., and Fujimura, O. (**1966**). ''Model for specification of the vocal-tract area function,'' J. Acoust. Soc. Am. **40**, 1271.

Dang, J., and Honda, K. (**1998**). ''Speech production of vowel sequences using a physiological articulatory model,'' in *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by Robert H. Mannell and Jordi Robert-Ribes, Vol. 5 (Sydney, Australia R.H.), pp. 1767–1770.

Demolin, D., Giovanni, A., Hassid, S., Heim, C., Lecuit, V., and Soquet, A. (**1997**). ''Direct and indirect measurements of subglottic pressure,'' *Proceedings of Larynx 97* (Marseille, France), pp. 69–72.

Djéradi, A., Guérin, B., Badin, P., and Perrier, P. (**1991**). ''Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method,'' J. Phonetics **19**, 387–395.

Edwards, J., and Harris, K. S. (**1990**). ''Rotation and translation of the jaw during speech,'' J. Speech Hear. Res. **33**, 550–562.

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 Vol. 3-4, pp. 11–38.

Fowler, C. A., and Saltzman, E. (**1993**). ''Coordination and coarticulation in speech production,'' Language and Speech **36**, 171–195.

Fromkin, V. A. (**1964**). ''Lip positions in American English vowels,'' Language and Speech **7**, 215–225 .

Gabioud, B. (**1994**). ''Articulatory models in speech synthesis,'' in *Funda-*

mentals of Speech Synthesis and Speech Recognition, edited by E. Keller (Wiley, Chichester), pp. 215–230.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). ''Factor analysis of tongue shape,'' J. Acoust. Soc. Am. **62**, 693–707.

Heinz, J. M., and Stevens, K. N. (**1965**). ''On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech,'' *Proceedings of the Fifth International Congress of Acoustics* (Liége, Belgium), Paper A44.

Hoole, P. (**1999**). ''On the lingual organization of the German vowel system,'' J. Acoust. Soc. Am. **106**, 1020–1032.

Hoole, P., and Kroos, C. (**1998**). ''Control of larynx height in vowel production,'' in *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by R. H. Mannell, J. Robert-Ribes, and E. Vatikiotis-Bateson (Australian Speech Science and Technology Association, Inc., Sydney, Australia), Vol. 2, pp. 531–534.

Hoole, P., Wismüller, A., Leisinger, G., Kroos, C., Geumann, A., and Inoue, M. (**2000**). ''Analysis of tongue configuration in multi-speaker, multi-volume MRI data,'' 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. 157–160.

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.

Laboissière, R., Ostry, D. J., and Feldman, A. G. (**1996**). ''Control of multi-muscle systems: Human jaw and hyoid movements,'' Biol. Cybern. **74**, 373–384.

Lallouache, M. T. (**1990**). ''Un poste Visage-Parole. Acquisition et traitement de contours labiaux [A ''Face-Speech'' workstation. Acquisition and processing of labial contours],'' *Proceedings of the 18th Journées d'Etude sur la Parole* (Montréal, Canada), pp. 282–286.

Liljencrants, J. (**1971**). ''A Fourier series description of the tongue profile,'' Speech Transmission Laboratory—Quarterly Progress Status Report Vol. 4/1971, pp. 9–18.

Lindblom, B. E. F. and Sundberg, J. E. F. (**1971**). ''Acoustical consequences of lip, tongue and jaw movements,'' J. Acoust. Soc. Am. **50**, 1166–1179.

Lindblom, B. E. F., Lubker, J., and Gay, T. (**1979**). ''Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation,'' J. Phonetics **7**, 141–161.

Maeda, S. (**1979**). ''Un modèle articulatoire de la langue avec des composantes linéaires,'' *Proceedings of the 10th Journées d'Etude sur la Parole* (Grenoble, France), pp. 152–163.

Maeda, S. (**1982**). ''A digital simulation method of the vocal tract system,'' Speech Commun. **1**, 199–299.

Maeda, S. (**1990**). ''Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,'' in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, The Netherlands), pp. 131–149.

Maeda, S. (**1991**). ''On articulatory and acoustic variabilities,'' J. Phonetics **19**, 321–331.

Matsuda, M., and Kasuya, H. (**1999**). ''Acoustic nature of the whisper,'' in *Proceedings of the 6th European Conference on Speech Communication and Technology*, Vol. 1 (Budapest, Hungary), pp. 133–136, September 1999.

Mawass, K., Badin, P., and Bailly, G. (**2000**). ''Synthesis of French fricatives by audio-video to articulatory inversion,'' Acta Acoustica **86**, 136–146.

Mermelstein, P. (**1973**). ''Articulatory model for the study of speech production'' J. Acoust. Soc. Am. **53**, 1070–1082.

Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). ''Two cross-linguistic factors underlying tongue shapes for vowels,'' J. Acoust. Soc. Am. **99**, 3707–3717.

Ostry, D., Vatikiotis-Bateson, E., and Gribble, P. (**1997**). ''An examination of the degrees of freedom of human jaw motion in speech and mastication,'' J. Acoust. Soc. Am. **40**, 1341–1351.

Payan, Y., and Perrier, P. (**1997**). ''Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis,'' Speech Commun. **22**, 185–205.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., Bailliet, H., Vescovi, C., and Castelli, E. (**1996**). ''Description of the flow through the vocal cords during phonation. Application to voiced sounds synthesis,'' Acta Acoustica **82**, 358–361.

Perkell, J. S. (**1974**). ''A physiological-oriented model of the tongue activity during speech production,'' Ph.D. dissertation, MIT, Cambridge.

Perkell, J. S. (**1991**). ''Models, theory and data in speech production,'' *Proceedings of the XIIth International Congress of Phonetic Sciences* (Université de Provence, Aix-en-Provence, France), Vol. 1, pp. 182–191.

Perrier, P., Ostry, D. J., and Laboissière, R. (**1996**). ''The equilibrium point hypothesis and its application to speech motor control,'' J. Speech, Language, and Hear. Res. **39**, 365–578.

Perrier, P., Payan, P., Perkell, J. S., Zandipour, M., Pelorson, X., Coisy, V., and Matthies, M. (**2000**). ''An attempt to simulate fluid-walls interactions during velar stops,'' 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. 149–152.

Revéret, L., and Benoît, C. (**1998**). ''A new 3D lip model for analysis and synthesis of lip motion in speech production,'' in *Proceedings of the International Conference on Auditory-Visual Speech Processing/Second ESCA ETRW on Auditory-Visual Speech*, edited by D. Burnham, J. Robert-Ribes, and E. Vatikiotis-Bateson (Terrigal-Sydney, Australia), pp. 207–212.

Sanguineti, V., Laboissière, R., and Ostry, D. J. (**1998**). ''A dynamic biomechanical model for neural control of speech production,'' J. Acoust. Soc. Am. **103**, 1615–1627.

Scully, C. (**1991**). ''The representation in models of what speakers know,'' in *Proceedings of the XIIth International Congress of Phonetic Sciences* (Université de Provence, Aix-en-Provence, France), Vol. 1, pp. 192–197.

Shadle, C. H., and Scully, C. (**1995**). ''An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences,'' J. Phonetics **23**, 53–66.

Sorokin, V. N., Gay, T., and Ewan, W. G. (**1980**). ''Some biomechanical correlates of jaw movements,'' J. Acoust. Soc. Am. Suppl. 1 **68**, S32.

Stark, J., Lindblom, B., and Sundberg, J. (**1996**). ''APEX: An articulatory synthesis model for experimental and computational studies of speech production,'' TMH-QPSR 2/1996, pp. 45–48.

Stetson, R. H. (**1928**). ''Motor phonetics. A study of speech movements in action,'' Archives Néerlandaises de Phonétique Expérimentale **3**, 216.

Stromberg, K, Scully, C., Badin, P., and Shadle, C. H. (**1994**). ''Aerodynamic patterns as indicators of articulation and acoustic sources for fricatives produced by different speakers,'' Institute of Acoustics **16**, 325–333.

Tiede, M. K. (**1996**). ''An MRI-based study of pharyngeal volume contrasts in Akan and English,'' J. Phonetics **24**, 399–421.

Vilain, A., Abry, C., and Badin, P. (**1998**). ''Coarticulation and degrees of freedom in the elaboration of a new articulatory plant: Gentiane,'' in *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by R. H. Mannell, J. Robert-Ribes, and, E. Vatikiotis-Bateson (Australian Speech Science and Technology Association, Inc., Sydney, Australia), Vol. 7, pp. 3147–3150.

Vilain, A., Abry, C., Badin P., and Brosda, S. (**1999**). ''From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modeling,'' 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 (Congress organizers at the Linguistics Department, University of California at Berkeley, San Francisco, CA), Vol. 3, pp. 2497–2500.

Westbury, J. R. (**1988**). ''Mandible and hyoid bone movements during speech,'' J. Speech Hear. Res. **31**, 405–416.

Westbury, J. R. (**1994**). ''On coordinate systems and the representation of articulatory movements,'' J. Acoust. Soc. Am. **95**, 2271–2273.

Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.

Zerling, J. P. (**1984**). ''Phénomènes de nasalité et de nasalisation vocaliques: Étude cinéradiographique pour deux locuteurs [in French],'' Trav. de l'Inst. de Phonétique de Strasbourg **16**, 241–266.