



**HAL**  
open science

## Modeling tongue-palate contact patterns in the production of speech

Noël Nguyen, Alain Marchal, Alain Content

► **To cite this version:**

Noël Nguyen, Alain Marchal, Alain Content. Modeling tongue-palate contact patterns in the production of speech. *Journal of Phonetics*, 1996, 24, pp.77-97. 10.1006/jpho.1996.0006 . hal-01393404

**HAL Id: hal-01393404**

**<https://hal.science/hal-01393404v1>**

Submitted on 7 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modeling Tongue-palate Contact Patterns in the Production of Speech

Noël Nguyen<sup>1</sup>, Alain Marchal<sup>2</sup>, and Alain Content<sup>1,3</sup>

*(1) Laboratoire de Psycholinguistique, FPSE, Université de Genève, 9 route de Drize, CH-1227 Genève, Suisse*

*(2) Laboratoire Parole et Langage, Université de Provence, 29 av. Robert-Schuman, F-13621 Aix-en-Provence, France*

*(3) Laboratoire de Psychologie expérimentale, Université Libre de Bruxelles, 117 av. Adolphe-Buyl, B-1050 Bruxelles, Belgique*

revised version submitted to the Journal of Phonetics

comments are welcome — please do not quote

running title: Modeling tongue-palate contact patterns

## Abstract

Previous studies on lingual movements in speech have led to the assumption that the tongue can be divided into a small number of functional blocks. This work was aimed at finding out whether such a subdivision into independent components may also be observed in patterns of tongue-palate contacts, as revealed by electropalatography (EPG). An attempt was made to reduce EPG data to a small number of articulatorily relevant parameters in an empirical way, and to model the configuration of the linguo-palatal contacts in speech as a combination of these parameters. Our corpus was composed of ten natural sentences recorded by two female native speakers of French. EPG data reduction was performed by means of a feedforward multilayer neural network on the one hand, and of a factor analysis on the other hand. The results showed that variations in the tongue-palate contact pattern mainly occurred along two dimensions, related to lingual contacts in *1)* the alveolar region and *2)* the palatal region, respectively. These results were consistent for both subjects, and provided evidence for the hypothesis that the tongue tip/blade and the tongue dorsum are two independently controllable articulators. In a second step, potential applications of this empirical model to studies of coarticulation and timing are illustrated, through a pilot investigation of /kl/ clusters. Finally, implications of our work for a comprehensive model of tongue movements are discussed.

## 1 Introduction

The tongue is characterized by a very high flexibility, as it does not have any internal bony skeleton. It is likely to have many different shapes, and has been described as “potentially [having] an enormous number of degrees of freedom in its movement” (Stone, 1991). However, in many studies, tongue movements in speech are modeled as a combination of a small set of elementary articulatory gestures. For example, Hardcastle (1976) considers that only four parameters are needed to describe the configuration of the tongue in the sagittal plane, namely the horizontal and vertical displacements of the tongue body and of the tip-blade. In Stone’s three-dimensional model (Stone, 1991), the tongue is divided into five lengthwise functional segments (anterior, middle, dorsal, posterior and root). Using a statistical method, Harshman et al. (1977) have demonstrated that the tongue sagittal shape can be reduced by means of a factor analysis to only two main components, namely the degree of raising/lowering of the front and of the back of the tongue, respectively (see also Maeda (1990), for a similar approach). In a recent study on the production of fricatives (Nguyen et al., 1994), high correlations between the  $x$  and  $y$  coordinates of three tongue points and the acoustic spectrum also tended to indicate that the tongue is controlled by a limited number of parameters.

This apparent limitation in the variety of actual tongue shapes and movements has been attributed to different factors. Obvious constraints are imposed by the basic structural properties of the tongue tissues, i.e. two closely situated points on the tongue cannot move independently of each other. Another important constraint is related to the fact that the tongue constitutes a hydrostat (a fluid-filled bag), and thus that its volume does not vary during displacement or deformation. Because of this volume incompressibility, “any change

in one dimension will cause a compensatory change in at least one other dimension” (Kier and Smith, 1985). To take an example, Sproat and Fujimura (1993) have suggested that, in the production of an alveolar lateral, the tongue blade narrowing in the cross-sectional plane is compensated for by an inversely proportional variation in the volume of the tongue along the front-back axis, i.e. an increase in the blade’s length, and that this biomechanical phenomenon is at the origin of the dorsum retraction which is observed for /l/. Finally, at the level of motor control, many studies have indicated that independent muscles may cooperate together so as to form single functional units, or “coordinative structures” (Fowler et al., 1980). Such coordinative structures necessarily involve a reduction in the number of degrees of freedom of each muscle. This could partly explain how a multidimensional dynamical system such as the vocal apparatus can be governed by means of a small set of articulatory parameters (Kelso et al., 1986).

While many models of tongue sagittal shape have been proposed, less work has been dedicated to the behavior of the tongue in the coronal plane. However, it seems possible to assert that tongue movements in that plane also have a limited number of degrees of freedom. According to Hardcastle (1976), such movements can be regarded as being controlled by three parameters (transverse cross-sectional configuration of the tongue body, degree of central grooving over the whole length of the tongue, and surface plane of the tongue dorsum). For Stone (1991), five semi-independent functional segments are involved in cross-sectional tongue movements (a medial segment and two symmetric lateral ones; see also Stone et al. (1992)). Although it is emphasized that the cross-sectional shape may be different at different locations along the sagittal axis, these data lead us to believe that movements of the tongue can be characterized, at a macroscopic level at least, in terms of a limited number of functional

units.

This work was aimed at building an empirical model of tongue movements on the basis of electropalatographic data. Electropalatography (EPG) is a technique for recording the timing and location of tongue contacts with the hard palate during speech (Marchal, 1988). Despite obvious limitations, it provides useful information about the configuration of the tongue in the coronal plane as well as in the sagittal plane. Our basic assumption was that the spatial arrangement of linguo-palatal contacts is governed by a small set of underlying parameters, that will be referred to as *control parameters* hereafter. In the present work, control parameters were extracted in an empirical way from various sets of EPG patterns, by means of statistical multidimensional methods.

## 2 Materials

Our data were extracted from the EUR-ACCOR multichannel database (Marchal and Hardcastle, 1993). The corpus consisted of one repetition of a set of ten French sentences (see Appendix). Sentences were preferred to the single lexical items and nonsense items also included in the database, since they had a larger variety of vowels and lingual consonants in various positions within words. It was assumed that materials of this kind would allow us to extract EPG control parameters which would not be specific to a particular phonemic sequence or gesture.

The subjects who participated in this study were two female native speakers of French (southern variety), AD (25 years), and GC (24 years), with no known speech or hearing disorders. The data were recorded at the Speech Research Laboratory at the University of Reading,

using the Reading EPG system (Hardcastle et al., 1989). The lingual contact patterns were sampled at a frequency of 200 Hz, and stored on a PC AT computer for further processing. The acoustic signal was recorded in parallel, low-pass filtered (cut-off frequency: 9 kHz), digitized (sample frequency: 20 kHz), and stored in a separate file. A detailed description of the recording procedure can be found in Marchal and Hardcastle (1993).

As shown in Figure 1, the Reading artificial palate has 62 electrodes arranged in eight rows and eight columns. Rows were numbered from 1 through 8 from front to back. Columns were numbered from 1 through 8 from left to right. A subdivision of the artificial palate into anatomically-based articulatory zones (Recasens, 1990) is also presented. Two main zones were distinguished along the sagittal dimension, namely the alveolar zone (rows 1–4) and the palatal zone (rows 5–8). Each of these two zones was itself subdivided into different subzones.

— insert Figure 1 about here —

In order to somewhat reduce the statistical redundancy in our materials and to make the computations less time-consuming, we removed from the EPG files the patterns which either had fewer than two activated electrodes, or were identical to the immediately preceding one on the time axis. Not surprisingly, dynamic variations of the linguopalatal contacts were in most cases rather slow, and long time series of identical EPG patterns were frequently observed. Such series would not have provided us with much more information than a single pattern as regards the spatial configuration of contacts, and would furthermore have had the disadvantage of giving more weight in the statistical analyses to articulatory stable states, to the detriment of rapid movements. In all, about 37 % of the EPG patterns were kept, out of nearly 12000 patterns. More precisely, the data were composed of 2524 palatograms

for speaker GC, and of 1906 palatograms for speaker AD. The removed patterns mainly corresponded to phonetic segments for which there is little or no contact between the tongue and the artificial palate (low vowel [a], high back vowel [u], labial and velar stops). In addition, frames situated in the stable part of coronal fricatives ([s] and [ʃ]) have also been removed.

### 3 Method

Our method for extracting EPG control parameters is schematized in Figure 2. It was, basically, a data reduction technique. In other words, a mathematical transform was employed to convert each pattern, as characterized by the “on” or “off” values of the 62 electrodes, into a smaller number of EPG parameters.

— insert Figure 2 about here —

In this framework, the main issue is to find a transform which guarantees the highest compression rate, i.e. which minimizes both the loss of information and the number of extracted parameters. The compression performance was determined here by resynthesizing each EPG pattern from the corresponding parameter values, and by computing the distance between the original pattern and the resynthesized one. Thus, in the example presented in Figure 2, it appears that a small amount of information was lost in the compression process, as two of the “on” electrodes in the original pattern (row 3) were switched off in the resynthesized pattern.



### 3.1 Preliminary analysis of the correlations between linguopalatal contacts in different areas

EPG control parameters were extracted from the EPG data using two different tools: a multilayer feedforward neural network on the one hand, and a principal-components analysis on the other hand. In a first step however, we attempted to determine to what extent the 62 electrode states of activations were inter-correlated. This preliminary analysis was motivated by the fact that EPG data reduction would not be possible if there was no relationship between the linguopalatal contacts at two different points on the palate. For the sake of simplicity, the EPG patterns were first reduced to vectors of eight values by summing the number of activated electrodes in each row across the artificial palate. A matrix of correlations was then calculated from this series of vectors. The results are shown in Table 1 for speaker GC, and in Table 2 for speaker AD. Row 1 is the most anterior one and row 8 the most posterior one.

— insert Table 1 about here —

— insert Table 2 about here —

As expected, these matrices show that the number of contacts touched in different areas of the artificial palate were not independent of each other. In fact, most of the correlation coefficients were statistically significant. Note, however, that these correlations tended to be higher for two adjacent rows than for more distant ones, although this was not entirely true for speaker AD, since in her case, the most anterior and posterior rows were more closely correlated with one another than with intermediate ones (for example,  $r(row8, row1) = .24$ , while  $r(row8, row2) = -.02$ ). In any case, the correlation coefficients given in Tables 1 and 2 were on average high enough to demonstrate that reducing the 62 electrode states of

activation to a smaller number of parameters was feasible.

### 3.2 Multilayer feedforward neural network

The neural network used in this work was a standard feedforward autoassociator, with one hidden layer (see Lippmann (1987), for a general tutorial on neural networks). The problem that the network had to solve was to reproduce, at the output layer, the EPG pattern presented to the input layer (*identity mapping*). As the hidden layer contained many fewer units than the input and output ones, the network was constrained to compress the data efficiently in order to perform the task. The pattern of activation of the hidden units thus constituted a compressed representation of the input. We hypothesized that the number of hidden units could be related to the intrinsic dimensionality of the data set.

The general architecture of the network is presented in Figure 3. A correspondence can be easily established between this figure and the three stages of the EPG compression/resynthesis process as depicted in Figure 2. Thus, the control parameters can be associated with the hidden unit states of activation. Parameter extraction is achieved by mapping the input pattern onto the hidden layer, and pattern resynthesis by mapping the hidden layer onto the output layer.

— insert Figure 3 about here —

In the construction of this network, the main objective was to find the optimal number of hidden units, i.e. the number had to be kept as small as possible whilst maintaining a high level of accuracy. This number could not be determined *a priori*, as our ultimate goal was to estimate the *intrinsic* dimensionality of the EPG data. Several researchers have proposed methods

for reducing the size of the hidden layer to a minimum (e.g. DeMers and Cottrell (1993); Fahlmann and Lebiere (1990); Oja (1991)). In the present work, we used a simple variant of back-propagation learning based on a sequential recruitment of the hidden units. Hidden units were introduced into the network one-by-one, and the connections from the input layer and to the output layer were iteratively modified during a fixed number of trials (100 cycles throughout the data set). Then, the connection weights to and from that particular unit were frozen, and a new hidden unit with full connections from the input layer and to the output layer was recruited. Two different networks were built in this way, one for each speaker. The learning rate was set to .01; no momentum was used. The computations were done using Tlearn, a program developed by J. Elman at the University of California at San Diego. This program implements the back-propagation learning algorithm as described in Rumelhart et al. (1986).

### **3.3 Factor analysis**

In a second step, and for each subject, the EPG patterns were recoded as vectors of 62 binary values (1: electrode “on”; 0: electrode “off”). These vectors were then subjected to a standard principal-components analysis (PCA). The resulting factor pattern matrix was rotated so as to either maximize or minimize the coefficient corresponding to each variable for each factor (Harman, 1976). The aim of this transformation was to have factors which are, as far as possible, in a one-to-one correspondence with groups of highly intercorrelated variables. It was assumed that the rotated factors would more clearly reflect local spatial correlations, such as those hypothesized by Stone (1991) for example (inasmuch as such correlations did really exist in our data and were stronger than nonlocal ones).

Just as autoassociative neural networks, factor analysis can be considered as a particular case of the data compression principle described in the introduction of the present section (see Figure 2). Thus, factors are formally equivalent to what we have called control parameters. They are extracted from EPG data by means of a (linear) transform the goal of which is to store as much information as possible into the lowest possible number of dimensions. Moreover, EPG patterns can be resynthesized from the factors, using another transform which directly derives from the first one<sup>1</sup>.

In fact, as shown by Baldi et al. (1989), autoassociative networks are functionally very similar to factor analysis. However, the latter is submitted to several constraints (linearity, orthogonality of factors) which make it generally less powerful than the former. It may also be hypothesized that a neural network is more specifically appropriate to EPG data than PCA, given the particular characteristics of this kind of data. Palatograms can indeed be compared to bidimensional images (with eight rows and eight columns) containing severe discontinuities between black and white dots. These discontinuities are reflected by the presence of high spatial frequencies with a relatively high amplitude in the image spectral domain (Nguyen, 1995). As a result, the images should be more difficult to reduce to a small number of factors than “smoother” greyscale images, for which the energy would mainly reside in the lowest spatial frequencies. EPG data reduction should be easier with a neural network, as the network units tend to be systematically either on or off, by virtue of the nonlinear (sigmoid) nature of their transfer function.

## 4 Results

### 4.1 Efficiency of the EPG data reduction

In the first place, an attempt was made to compare the respective efficiency of the two methods in reconstructing the original EPG patterns from a small set of underlying parameters. The reconstructed EPG patterns were converted into black and white images by rounding to the nearest integer the values of the 62 picture elements, which were originally continuous and comprised between 0 and 1. Then an error index was defined, by computing the number of electrodes for which the estimated value was wrong (1 instead of 0 and *vice-versa*), in each reconstructed pattern. The overall mean error as a function of the number of parameters is shown in Figure 4, for both methods and both speakers.

— insert Figure 4 about here —

As expected, the mean error diminished in both methods as the number of parameters was increased. However, Figure 4 indicates that the performance of the neural net was systematically better than that of the factor model. Paired t-tests<sup>2</sup> showed that the error was significantly lower for the former than for the latter, for any given number of parameters, and for both subjects ( $p < .001$  in all cases). A difference was also found between the two subjects, the mean error being lower for GC. Interestingly, it appeared that the mean error rapidly reached a value which could be considered as being rather low when it was related to the EPG data *overall variability*. In the present study, this variability was estimated by calculating, for each pattern, the number of electrodes having a different value in comparison with the speaker-specific mean pattern. The average variability was 12.6 for subject GC and 11.5 for subject AD.

Pairwise correlations between the factors and the hidden unit states of activation were also computed, and are presented in Table 3. This table reveals two interesting trends. In the first place, there was a close correspondence between the main factors and hidden units. For speaker GC for example, the first hidden unit was highly and almost exclusively correlated with factor 1 ( $r = .98$ ), and the second hidden unit with factor 2 ( $r = .90$ ). This shows that the parameters extracted first from the EPG data by the network, were most likely similar to principal components. Note, however, that hidden unit 2 was also correlated with factor 3 as well as factor 2 for speaker AD.

— insert Table 3 about here —

It is also noticeable that the hidden unit states of activation were, just as principal components, almost independent of each other. This suggests that the network internal representation of the input was *local* rather than distributed over several hidden units. Therefore, there is a strong possibility that each hidden unit was associated with a specific set of features in the input pattern. This hypothesis is discussed in more detail in the next section.

## 4.2 Articulatory significance of the EPG control parameters

As the neural network hidden units appeared to correspond well to the factors obtained in the PCA, while allowing a more accurate resynthesis of the EPG patterns, the following discussion will be restricted to the network analysis.

The most important issue was whether it was possible to interpret in articulatory terms the control parameters that had been extracted from the EPG data by means of the network. This issue was addressed in two different ways. First, a relationship was established between these

parameters and the original EPG data, namely the 62 electrode “on” or “off” values. Second, we attempted to synthesize “articulatory primitives” by manipulating the control parameters independently of each other, using a method similar to that of Harshman et al. (1977) and Maeda (1990), in their work on modeling the vocal-tract sagittal shape.

The correlation coefficients between the activation states of the first four hidden units and the electrode values are presented in Figure 5, as three-dimensional plots. In each of these plots, the abscissa and the ordinate represent the coronal and front-back dimensions, respectively. Thus, subdivisions along the first axis refer to the palate columns, whereas subdivisions along the second one refer to the palate rows. The  $z$  axis is a correlation scale. Strong relationships between a hidden unit and a specific palatal zone are represented either by a peak (positive correlations), or by a trough (negative correlations), on the corresponding correlation surface<sup>3</sup>.

— insert Figure 5 about here —

Figure 5 shows that, for speaker GC, the first hidden unit was mainly associated with electrodes located in the alveolar zone. High positive correlation values ( $> .7$ ) were more particularly observed on both sides (columns 1–3 and 6–8) of the anterior part of the palate. The degree of correlation decreased as the electrodes were closer to the midline or to the palate posterior edge. Conversely, hidden unit 2 was chiefly correlated with electrodes situated along the lateral edges (columns 2–3 and 6–7) in the palatal zone, as revealed by the presence of two correlation peaks in this area, whereas a gradual decline of the correlation values was observed from back to front. Relating the third hidden unit to a specific anatomical region was more difficult. However, the highest correlation values appeared to correspond to electrodes which

were all located in the post-alveolar (rows 3–4) and in the prepalatal zones (row 5). Finally, the fourth hidden unit was characterized by positive correlations with electrodes close the midline and in the rear part of the palate (postpalatal zone), and by negative correlations with electrodes along the lateral edges.

Very similar results were found for speaker AD. The main difference resided in the fact that hidden unit 1 was not as closely correlated with the frontmost electrodes (front alveolar zone) as for speaker GC. Note also the presence of troughs (reflecting negative correlations) on the correlation surface at the level of the backmost palatal row, for the third hidden unit.

It may therefore be said that the first two hidden units were related to *specific palate zones*, which were identical for both speakers. Thus, an increase in the activation of hidden unit 1 was associated with a higher number of contacts in the alveolar region. In the same way, an increase in the activation of hidden unit 2 was accompanied by a rise in the number of contacts in the palatal region. It might also be possible to establish a correspondence between hidden unit 3 and the postalveolar zone, as between hidden unit 4 and the medial part of the palate. However, note that the degree of correlation with the corresponding electrodes never exceeded .7, for these two hidden units<sup>4</sup>.

These basic trends were also partially encountered in the “EPG primitives” respectively corresponding to hidden units 1 and 2, and which are shown in Figure 6, for both speakers. These primitives were synthesized by driving the network output layer with each hidden unit separately. The hidden unit activation level was increased from 0 to 1 in five steps. In the present study, a tongue-palate contact pattern was viewed as a combination of such primitives, which were similar in that sense to the “elementary gestures” in Maeda’s model (1990) for example. In the following, hidden units 1 and 2 will be referred to as HU1 and



HU2, respectively.

— insert Figure 6 about here —

As shown in Figure 6, the EPG primitive associated with HU1 may be regarded as a raising of the tongue tip-and-blade, resulting in linguopalatal contacts in the front alveolar and postalveolar regions. Similarly, it may be assumed that the primitive corresponding to HU2 represents an upward movement of the tongue dorsum, with linguopalatal contacts in the palatal and postpalatal zones. In both cases, the initial (leftmost) pattern was characterized by contacts on each side of the palate. These results were fairly consistent across our two speakers. Note, however, that the dynamic range of such synthetic gestures was rather limited. In particular, it appeared that modeling an alveolar or postpalatal *closure* was apparently not possible, at least on the basis of the first two hidden units.

In summary, our results lead us to believe that tongue-palate contact patterns can be divided into a small number of (semi)independent “building blocks”. The two most important of these building blocks were closely associated with 1) the alveolar zone and 2) the palatal zone. Such a subdivision of the electropalate into two different zones along the front/back dimension proved to be rather robust since it was observed for both speakers. By performing several trials with different sets of initial weights, we also controlled that the data were encoded in a way which did not depend on the network’s starting configuration.

### 4.3 Application to the study of /kl/ clusters

In this section, we will attempt to demonstrate that dividing EPG patterns into independent “building blocks” can provide useful clues on issues such as coarticulation and articulatory

timing. This preliminary analysis will be focused on /kl/ clusters. The production of clusters of this type is interesting in that it involves a coordination of the tongue tip/blade system and of the tongue body system. In previous studies (Hardcastle, 1985; Gibbon et al., 1993), /kl/ clusters were used to examine the temporal and spatial characteristics of such inter-articulator coordination. In many cases, a temporal overlap in the articulatory gestures for /k/ and /l/ was observed, i.e. the onset of /l/ closure began simultaneously, or even before, the release of the /k/ closure. It was found that the amount of overlap was dependent on the rate of utterance, on the type of /k-/l/ junctural boundary (Hardcastle, 1985), and on language-dependent differences in degree of velarization for /l/ (Gibbon et al., 1993). We hypothesize that such a tendency to freely coarticulate the anterior with the posterior part of the tongue should be also encountered in the dynamic variations of the first two EPG control parameters, inasmuch as these parameters are related to the tip/blade system and to the tongue body system, respectively.

For this part of our study, the corpus was made up of /kl/ clusters combined with /a/ and /i/ in both symmetrical and asymmetrical /VklV/ sequences. These sequences were repeated ten times each, in a random order. They were recorded in the same conditions as the material presented above, and by the same speakers (GC and AD). The EPG and audio data are also part of the EUR-ACCOR database, and they were already used in a cross-linguistic study on lingual coarticulation in /kl/ sequences carried out by Gibbon et al. (1993). For each speaker and each EPG pattern, the hidden unit states of activation were calculated using the matrix of weights obtained at the end of the preceding training phase (section 3.2).

The temporal variations of the activation level for HU1 and HU2 in the production of /akla/ by speaker GC are shown in Figure 7. Cubic spline functions were employed for smoothing

both curves (see Keller and Ostry (1983)). Also presented are palatograms associated with important articulatory events, as defined in Gibbon et al. (1993), namely, from left to right: 1) the onset of the /k/ closure, 2) the release of the /k/ closure, 3) the approach to the alveolar closure for the lateral <sup>5</sup>, and 4) the release of the lateral. There was clearly a close correspondence between these articulatory events and the acoustic signal (shown in the upper panel).

— insert Figure 7 about here —

Figure 7 indicates that HU2 approximately reflected the variations of the linguopalatal contacts in the palatal region. Thus, the leftmost palatogram (which was the first one to show a complete postpalatal closure in the EPG signal) coincided with the beginning of a plateau in the HU2 activation level and, reciprocally, the release of the stop closure was associated with a decline in this level. A relationship may be established in the same way between HU1 and contacts in the first four rows of the electropalate, as the onset and offset of the lateral, as seen on the EPG trace, roughly coincided with rapid changes in HU1 state of activation. Although dynamic variations of the hidden units only partially corresponded to the articulatory events identified, these data again suggest that HU1 and HU2 may be respectively associated with bracing against the palate of the tongue tip-and-blade and of the tongue dorsum. The temporal variations of the hidden unit states of activation averaged over ten repetitions for each /VklV/ sequence and each speaker are displayed in Figure 8.

— insert Figure 8 about here —

Several interesting trends are brought out for both speakers in this figure. In the first place, the two parameters appeared to be sensitive to the influence of the surrounding vowels. This

was especially true for HU2, whose maximum value could double in the vicinity of /i/. Such variations are not surprising, given the well-known susceptibility of the place of articulation for /k/ to the vowel context. A rather stable temporal relationship was also observed between HU1 and HU2 activation curves. Thus, the onset of HU2 activation decline (close to the release of the velar closure) coincided with a local minimum in the HU1 activation level, as we would expect, given the fact that /k/ was most obviously produced with little or no lingual contact in the anterior part of the palate. Reciprocally, HU1 maximum value (close to the end of the lateral) was associated with a local minimum on the HU2 activation curve. This may be taken as an indication that the tongue dorsum was actively lowered in the production of /l/.

The influence of vowels on consonants was quantitatively assessed by measuring HU1 and HU2 activation values at the stop release and at the lateral release (as defined above), for each item and each speaker. For each of these two points, we computed the mean difference in the activation value of each hidden unit as a function of 1) the first vowel, and 2) of the second vowel. This measure was somewhat similar to the “coarticulation index” proposed by Farnetani et al. (1989). The results are shown in Table 4.

— insert Table 4 about here —

As regards HU1, Table 4 shows that the activation level was sensitive to the influence of the first vowel at the stop release (for speaker GC), and of the second vowel at the lateral release (for both speaker GC and AD). The activation level of HU2 significantly varied as a function of the first as well as of the second vowel at the stop release, and of the second vowel at the lateral release, for both speakers. To put it differently, our results indicated that

contextual variations of the tongue-palate contacts in the anterior part of the palate were mainly observed at the release of the lateral, and were attributable to the influence of the following vowel. Conversely, the tongue-palate contact pattern in the posterior part of the palate appeared to be sensitive to the following vowel throughout the consonant cluster, and to the preceding vowel at the release of the velar.

Differences in the magnitude of the vowel-dependent variations between HU1 and HU2 may be related to the functional distinction which is often established between the tongue tip-and-blade and the tongue body (Öhman, 1966; Perkell, 1969). On the one hand, the tongue body is involved in the production of the vowels as in that of dorsal consonants, and therefore was raised for /k/ as well as for /l/ in the vicinity of an /i/, due to the coarticulatory influence of the vowel. On the other hand, the tongue tip-and-blade is essentially active in the production of coronal consonants, and showed less variations in its position, according to the vowel context. It should also be emphasized that the different parts of the tongue are evidently submitted to different constraints in the production of the velar stop and of the alveolar lateral. Thus, /k/ may well have “blocked” the potential influence of the first vowel on /l/, since the tongue body intervenes on the formation of the velar closure, and the tip-and-blade is constrained at the same time not to contact the roof of the oral cavity. On the contrary, /l/ may have allowed the second vowel to be prepared as soon as the release of /k/, since the tongue body can move rather freely for the lateral as a function of the vocalic context.

An attempt was also made to characterize the temporal coordination of the /k/ and /l/ gestures (as reflected in the hidden unit activation curves) in quantitative terms. Our analysis

was focused on the temporal relationship between the onset of the decline in HU2 activation level, and the point where HU1 reaches its maximum value, over the consonant cluster. Rather than measuring the time interval between these two events (which might depend on the overall tempo, see Hardcastle (1985)), we tried to determine whether HU1 maximum value occurred *at a specific point* along the HU2 declining activation curve. For this purpose, a phase angle analysis was performed, according to the method described in Kelso et al. (1986). To be more specific, we computed the angle at which HU1 maximum value was reached relative to HU2 onset of activation decline, in the second hidden unit phase plane. Note that the phase angle was only used as a convenient statistical index for describing the relationship between two temporal variables, and that it did not involve any specific assumption as regards the dynamical organization of articulatory gestures. In accordance with Kelso et al.'s mathematical definition, the phase angle increased as the HU1 activation peak was located further on the HU2 phase plane trajectory. For example, a phase angle higher than 180 degrees indicated that this peak was posterior to a local minimum in the HU2 level of activation. The mean phase angles for each speaker and each item are shown in Figure 9.

— insert Figure 9 about here —

Figure 9 reveals that the phase angle was rather stable over the four items for speaker AD. No significant difference was found as a function of vowel 1 or vowel 2. Inversely, clear variations in the phase angle in different vowel contexts were observed for speaker GC. An ANOVA showed that the phase angle was significantly higher when the following vowel was an /i/ rather than an /a/ ( $F(1,35) = 65.68, p < .0001$ ). Differences according to the preceding vowel appeared also to be significant, although to a lesser degree ( $F(1,35) = 4.1, p = .05$ ),

and there was a significant interaction between vowel 1 and 2 ( $F(1, 35) = 17.54, p < .0002$ ). Couched in articulatory terms, these results lead us to believe that, for speaker AD, the relative timing of the dorsum and tip-and-blade maximum extent of movement, in /kl/ clusters, remained rather stable across changes in the vowel context. For speaker GC on the contrary, the stable phase and offset of the lateral appeared to be “delayed” with respect to the release of the velar stop in the /-i/ context, as opposed to the /-a/ context, this delay being substantially larger when the preceding vowel was /a/ than when it was /i/. Thus, it seems that for GC, there was a significant difference in the orchestration of the articulatory gestures for /k/ and /l/ over time, depending on whether the following vowel was /i/ or /a/, particularly when V1 was /a/. Note also that variations in the phase angle were larger in front of /i/ than in front of /a/, for both speakers.

The existence of this vowel-specific articulatory strategy, as captured by the phase angle analysis, was confirmed by visual inspection of the EPG signal itself. Figure 10 shows for example an EPG printout of the transition between /k/ and /l/ in the production of /akla/ (upper panel) and /akli/ (lower panel) by speaker GC. In both sequences, the first frame was the one which immediately preceded the release of the /k/ closure.

— insert Figure 10 about here —

Figure 10 demonstrates that the /k/-to-/l/ articulatory transition was clearly different according to the following vowel. For /akla/, the tongue-palate contact pattern remained relatively stable over an interval of 70 ms (from the release of the /k/), then changed abruptly at frame 60, which showed a sudden increase of contacts in the first four rows of the palate (coinciding with the onset of the lateral). Inversely, for /akli/, the transition between /k/

and /l/ appeared to be much smoother, as the /l/ onset (frame 65) was preceded by a gradual progression of the contacts along the lateral sides of the palate. In other words, when the subsequent vowel was an /a/, the /k/-to-/l/ transition appeared to be achieved by *a*) lowering the dorsum at the release of the closure for /k/ and *b*) rapidly raising the tip-and-blade toward the alveolar and postalveolar region. When V2 was an /i/ on the contrary, the dorsum was most likely maintained in a high position and the contacts between the tongue and the palate were progressively extended forward.

## 5 Discussion

To summarize, this work showed that EPG tongue-palate contact patterns may be reduced to a small number of independent parameters, by means either of a multilayer feedforward neural network, or of a factor analysis. Consistent results were obtained for both speakers. The two main control parameters appeared to be associated with linguopalatal contacts in the alveolar zone and in the palatal zone, respectively. If we assume that the electropalate alveolar region corresponds approximately to the tongue tip-and-blade, and that the palatal region corresponds to the tongue dorsum (Recasens, 1990), then our data provide evidence for the hypothesis according to which the tongue tip/blade system, and the tongue dorsum system, are two independently controllable articulators (Hardcastle, 1976).

These data also partially confirm the model of three-dimensional tongue movements proposed by Stone (1991). In this model, the front part of the tongue is subdivided into three sagittal segments, namely the anterior segment, the middle segment and the dorsal segment, involved in the production of alveolars, post-alveolars and velars, respectively. However, note that



only limited evidence was found for an independent middle segment. Although the network third hidden unit appeared to be associated with the postalveolar region for both speakers (see Figure 5), this relationship was not strong enough to lead us to believe that the middle segment could contact the roof of the oral cavity independently of the frontmost part of the tongue. Most obviously, such results should be considered as being specific to French. Stone has observed that the tongue middle section is raised independently of both the anterior and dorsal segments in the production of the American English bunched [r] (Stone, personal communication). If our analyses were to be extended to phonetic units of this kind, it may be assumed that a third EPG “building block” specifically corresponding to the middle part of the tongue would emerge.

Note also that we found no clearly identifiable EPG lateral primitive. In other words, the tongue-palate contact configuration along the coronal dimension did not appear to be independent of that along the front-back dimension (although the network fourth hidden unit tended to be associated with the former only, for both speakers). Although this might be taken as an indication that there is no independent control of the tongue movements in the coronal plane, a post-hoc analysis revealed that our negative finding may also be attributed to a bias in the electrode relative frequencies of activation. For the great majority of EPG patterns used in the first part of the analyses (sentences), contacts between the tongue and the lateral edges of the palate were observed. More specifically, the electrodes located on columns 1 and 8 were activated in more than 70% of cases<sup>6</sup>, for both speakers. Such a high frequency of lateral contacts can hardly be considered as reflecting the articulatory characteristics of some particular phonetic unit in our material, and it may well be due to a basic anatomical factor, i.e. the vaulted shape of the palate. It can indeed be assumed that, be-

cause of this vaulted shape, lateral contacts were often established as soon as the tongue, and particularly the dorsum, touched the palate. This would explain why there are lateral contacts for both the alveolar and palatal EPG primitives (Figure 6). Furthermore, this would explain the fact that low correlations were generally found between the network hidden units and the electrodes along the lateral edges, particularly in the back part of the palate (Figure 5). In other words, the absence of any lateral primitive could result from the fact that the configuration of the linguopalatal contacts along the coronal dimension was not a particularly salient feature for the network, as there were few variations in the number of contacts in the leftmost and rightmost columns<sup>7</sup>.

As indicated in the introduction of the paper, the approach used in the present work was an empirical one, in that EPG control parameters were defined on the basis of observed correlations in the spatial arrangement of tongue-palate contacts. It was hypothesized that the methods employed by other researchers for modelling the vocal tract sagittal shape (Harshman et al., 1977; Maeda, 1990) could be extended to EPG patterns, and that EPG control parameters could be extracted in a “bottom-up” fashion from a set of empirical data. Most obviously, empirical models of this kind do not allow us to *explain* how the motor system can succeed in governing the articulatory movements by means of a small number of underlying parameters. In that sense, control parameters as defined above have to be viewed as having a descriptive value only. Further theoretical and experimental investigations on the dynamics of speech movements are needed to determine whether these parameters have a significance at the level of motor programming.

## Acknowledgements

This work was supported by project ESPRIT ACCOR BRA # 3729, ESPRIT Working Group # 7098, and Swiss project OFES # 93.0351. We are grateful to Jeff Elman for making the Tlearn software freely available to us. Special thanks are due to Fiona Gibbon for very useful discussions. Comments from Paul Dagenais, Alice Faber, Ruth Kearns, John Kingston, Daniel Recasens, and Maureen Stone also led to major improvements, and are very gratefully acknowledged.

## References

- Baldi, P., and Hornik, K. (1989) Neural networks and principal component analysis: learning from examples without local minima, *Neural Networks*, **2**, 53–58.
- Clarke, R.J. (1985) *Transform Coding of Images*. London: Academic Press.
- DeMers, D., and Cottrell, G.W. (1993) Nonlinear dimensionality reduction. In *Advances in Neural Information Processing Systems 5* (S.J. Hanson, J.D. Cowan, and C.L. Giles, editors), pp. 580–587. San Mateo, California: Morgan Kaufmann.
- Fahlmann, S.E., and Lebiere, C. (1990) The Cascade-Correlation learning architecture. In *Advances in Neural Information Processing Systems 2* (D. Touretzky, editor), pp. 524–532. Mountain View, CA: Morgan Kaufmann.
- Farnetani, E., Hardcastle, W.J., and Marchal, A. (1989) Cross-language investigation of lingual coarticulatory processes using EPG. In *Proceedings of EUROSPEECH '89*, Paris, Sept. 89 (J.-P. Tubach and J.-J. Mariani, editors), pp. 429–432.
- Fowler, C.A., Rubin, P., Remez, R.E., and Turvey, M.T. (1980) Implications for speech production of a general theory of action. In *Language Production, vol. 1: Speech and Talk* (B. Butterworth, editor), pp. 373–420. London: Academic Press.
- Gibbon, F., Hardcastle, W.J., and Nicolaidis, K. (1993) Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation, *Language and Speech*, **36**, 261–278.
- Hardcastle, W.J. (1976) *Physiology of Speech Production. An Introduction for Speech Scientists*. London: Academic Press.

- Harcastle, W.J. (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences, *Speech Communication*, **4**, 247–263.
- Hardcastle, W.J., Jones, W., Knight, C., Trudgeon, A., and Calder, G. (1989) New developments in electropalatography: a state-of-the-art report, *Clinical Linguistics & Phonetics*, **3**, 1–38.
- Harman, H.H. (1976) *Modern Factor Analysis*. Chicago: The University of Chicago Press.
- Harshman, R., Ladefoged, P., and Goldstein, L. (1977) Factor analysis of tongue shapes, *Journal of the Acoustical Society of America*, **62**, 693–707.
- Keller, E., and Ostry, D.J. (1983) Computerized measurement of tongue dorsum movements with pulsed echo ultrasound, *Journal of the Acoustical Society of America*, **73**, 1309–1315.
- Kelso, J.A.S., Saltzman, E.L., and Tuller, B. (1986) The dynamical perspective on speech production: data and theory, *Journal of Phonetics*, **14**, 29–59.
- Kier, W.M., and Smith, K.K. (1985) Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats, *Zoological Journal of the Linnean Society*, **83**, 307–324.
- Lippmann, R.P. (1987) An introduction to computing with neural nets, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-4(2)**, 4–22.
- 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 Modelling* (A. Marchal and W.J. Hardcastle, editors), pp. 131–149. Dordrecht: Kluwer.
- Marchal, A. (1988). *La Palatographie*. Paris: CNRS.
- Marchal, A., and Hardcastle, W.J. (1993) ACCOR: instrumentation and database for the cross-language study of coarticulation, *Language and Speech*, **36**, 137–153.
- Nguyen, N. (1995) EPG bidimensional data reduction, *European Journal of Disorders of Communication*, **30**, in press.
- Nguyen, N., Hoole, P., and Marchal, A. (1994) Regenerating the spectral shape of [s] and [ʃ] from a limited set of articulatory parameters, *Journal of the Acoustical Society of America*, **96**, 33–39.
- Öhman, S.E.G (1966) Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America*, **39**, 151–168.
- Perkell, J.S. (1969) *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, Mass.: M.I.T. Press.
- Recasens, D. (1990) The articulatory characteristics of palatal consonants, *Journal of Phonetics*, **18**, 267–280.
- Oja, E. (1991) Data compression, feature extraction, and autoassociation in feedforward neural networks. In *Artificial Neural Networks* (T. Kohonen, O. Simula, and J. Kangas, editors), pp. 737–745. North-Holland: Elsevier.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning internal representations by error propagation. In *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, Vol. 1 (D. Rumelhart and J.L. McClelland, editors), pp. 318–362. Cambridge, Mass.: M.I.T. Press.

Sproat, R., and Fujimura, O. (1993) Allophonic variation in English /l/ and its implications for phonetic implementation, *Journal of Phonetics*, **21**, 292–311.

Stone, M. (1991) Toward a model of three-dimensional tongue movement, *Journal of Phonetics*, **19**, 309–320.

Stone, M., Faber, A., Raphael, L.J., and Shawker, T.H. (1992) Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [ʃ], and [l], *Journal of Phonetics*, **20**, 253–270.

Winer, B.J. (1971) *Statistical Principles in Experimental Design*. NY: McGraw-Hill.

## Appendix: list of the sentences used in the experiment

1. “La cousine de Vichy épousa un hippie à Toulouse” (“my cousin from Vichy married a hippy at Toulouse”).
2. “Le gala de Vichy appâta les fadas” (“the reception at Vichy attracted the crackpots”).
3. “Le toubib dépista tout à coup le choucas” (“the doc suddenly tracked down the jackdaw”).
4. “Le papou liquida la soupe au pistou” (“the Papuan finished off the pesto soup”).
5. “La toupie cassa le ciment de la datcha” (“the spinning-top broke the datcha’s cement”).
6. “Le loustic est assoupi près d’Annie” (“The chap is dozing close to Annie”).
7. “Jacques dit qu’il faut appeler un médecin” (“Jacques says that we have to call a doctor”).
8. “C’est maintenant que la smala les acclame” (“The tribe acclaims them now”).
9. “L’hôtesse cherche notre ticket de concert” (“the hostess is looking for our concert ticket”).
10. “Toutes les clefs sont refaites depuis hier” (“all the keys have been recut since yesterday”).

## Footnotes

<sup>1</sup>In fact, there is a direct relationship between the percentage of variance accounted for by a set of factors, and the accuracy of the pattern resynthesis achieved on the basis of these factors. In a certain sense, it is possible to say that pattern resynthesis is an intrinsic component of the factor analysis, in that factors are devised so as to minimize the distance between the resynthesized patterns and the original ones.

<sup>2</sup>Note that the distribution of errors was highly skewed to the left, with a majority of values close to zero. As paired t-tests are appropriate when the two variables follow a normal distribution only, error values were submitted to a preliminary transform, which was:

$$error' = 2 \times \arcsin(\sqrt{error})$$

This transform made the distribution of errors closer to a normal one (Winer, 1971).

<sup>3</sup>The most usual way of characterizing the information extracted by a given hidden unit from a set of data is to examine the weights of the connections between each input unit and this hidden unit. However, correlation analyses were judged more appropriate to the present paper, as they might be more meaningful to readers who are still unfamiliar with neural nets.

<sup>4</sup>An important question which arises here is how high a correlation coefficient should be for a hidden unit to be related to a given electrode. As most of the coefficients in our correlation surfaces appeared to be highly significant, it was impossible to use the significance level as a discriminant criterion. In our analyses, a correlation coefficient was somewhat arbitrarily considered as being “high” when it was greater than .7. As the exact limits of the region associated with each hidden unit on the palate were dependent on this threshold, they also have to be viewed as being, up to a certain extent, arbitrary.

<sup>5</sup>In Gibbon et al. (1993), the approach to alveolar closure for the lateral was taken as the first frame with any of the four midsagittal electrodes contacted in the first four rows.

<sup>6</sup>Average value over rows 2–8.

<sup>7</sup>Another factor may have to be taken into account, that is the width of the artificial palate. Noticeably, contacts along the lateral edges of the palate were observed in many occurrences of /l/ (see for example Figures 7 and 10). Thus, it is possible that the electrode array does not extend far enough laterally to display adequately the lowering of the lateral margins of the tongue in the production of /l/ (Stone, 1991; Stone et al., 1992). This may also explain why there was no independent EPG lateral primitive.



## Figure captions

1. Schematic representation of the Reading artificial palate.
2. Overview of the parameter extraction method.
3. Architecture of the neural network. Not all the connections shown.
4. Mean error in reconstructing the original EPG patterns, as a function of the number of parameters used in the factor model and in the neural net, for both speakers. The error represents the average number of electrodes in an EPG pattern whose estimated values were wrong.
5. Correlation coefficients between the hidden unit states of activation, and the 62 electrode on/off values, displayed as  $x$ - $y$ - $z$  plots. Results presented for the first four network hidden units, and for each speaker. In each plot, the  $x$  and  $y$  axes represent the coronal and the front-back dimensions, respectively. The  $z$  axis is a correlation scale.
6. EPG primitives associated with the network first two hidden units, for both speakers. Each primitive was generated by increasing the hidden unit activation level from 0 to 1 in five steps, whilst keeping the level of the other hidden units at 0. At each step, the network output vector was converted into a binary pattern by rounding each output unit value to the nearest integer.
7. Acoustic signal and hidden unit activation curves for the item /akla/ (speaker GC). Selected EPG patterns are also shown.
8. Hidden unit activation curves averaged over 10 repetitions, for each /VklV/ sequence,

and for both speakers. Dashdot line: hidden unit 1; solid line: hidden unit 2.

9. Phase angle at which HU1 maximum value was reached relative to HU2 onset of activation decline. Mean values and associated standard deviations for each speaker and each item.
10. EPG printout of the transition between /k/ and /l/ in /akla/ (upper panel) and /akli/ (lower panel) for speaker GC. The time interval between two adjacent patterns is 10 ms.

## Tables

Table 1: Pairwise correlations between the numbers of activated electrodes in the eight palate rows. Speaker GC. Row 1 is the most anterior one and row 8 the most posterior one. All the correlations are significant at the .05 level, except  $r(row8, row4)$ .

	row 1	row 2	row 3	row 4	row 5	row 6	row 7	row 8
row 1	1.00							
row 2	0.83	1.00						
row 3	0.58	0.79	1.00					
row 4	0.41	0.59	0.83	1.00				
row 5	0.25	0.43	0.58	0.78	1.00			
row 6	0.08	0.24	0.35	0.52	0.85	1.00		
row 7	-0.04	0.09	0.20	0.38	0.73	0.88	1.00	
row 8	-0.07	-0.05	-0.09	0.04	0.34	0.53	0.70	1.00

Table 2: Pairwise correlations between the numbers of activated electrodes in the eight palate rows. Speaker AD. All the correlations are significant at the .05 level, except  $r(row7, row2)$ ,  $r(row7, row3)$ , and  $r(row8, row2)$ .

	row 1	row 2	row 3	row 4	row 5	row 6	row 7	row 8
row 1	1.00							
row 2	0.66	1.00						
row 3	0.44	0.81	1.00					
row 4	0.36	0.56	0.81	1.00				
row 5	0.39	0.37	0.48	0.70	1.00			
row 6	0.30	0.20	0.25	0.46	0.82	1.00		
row 7	0.17	0.02	0.01	0.17	0.52	0.76	1.00	
row 8	0.24	-0.02	-0.14	-0.10	0.15	0.32	0.63	1.00

Table 3: Pairwise correlations between the first four factors and the first four hidden units for both speakers.

	speaker GC				speaker AD			
	HU1	HU2	HU3	HU4	HU1	HU2	HU3	HU4
HU1	1.00				1.00			
HU2	0.13	1.00			0.10	1.00		
HU3	0.15	0.08	1.00		0.33	-0.03	1.00	
HU4	-0.11	-0.03	0.03	1.00	0.27	0.00	-0.03	1.00
fac. 1	0.98	0.08	0.22	-0.11	0.95	0.03	0.44	0.34
fac. 2	0.27	0.90	0.21	0.04	0.13	0.60	0.11	0.38
fac. 3	-0.07	-0.26	0.46	0.51	-0.21	-0.68	0.14	0.44
fac. 4	-0.01	-0.11	0.59	-0.45	-0.02	-0.06	0.74	-0.36

Table 4: Mean difference in each hidden unit activation level at the stop release and at the lateral release, as a function of the first or second vowel, for each subject.  $\Delta V1$ : mean activation for  $V1 = /i/$  minus mean activation for  $V1 = /a/$ ;  $\Delta V2$ : mean activation for  $V2 = /i/$  minus mean activation for  $V2 = /a/$ . Differences which are significantly better than zero are marked: \* ( $p < .01$ ), \*\* ( $p < .001$ ).

	speaker GC		speaker AD	
	stop release	lateral release	stop release	lateral release
HU1	$\Delta V1: 0.09^{**}$	$\Delta V1: 0.00$	$\Delta V1: -.06$	$\Delta V1: -.02$
	$\Delta V2: 0.04$	$\Delta V2: 0.28^{**}$	$\Delta V2: 0.01$	$\Delta V2: 0.30^{**}$
HU2	$\Delta V1: 0.27^{**}$	$\Delta V1: 0.04$	$\Delta V1: 0.20^*$	$\Delta V1: 0.00$
	$\Delta V2: 0.18^*$	$\Delta V2: 0.37^{**}$	$\Delta V2: 0.20^*$	$\Delta V2: 0.59^{**}$