A study of the French vowels through the main constriction of the vocal tract using an acoustic-to-articulatory inversion method

Slim Ouni[*] and **Yves Laprie**[†]

[*] PSL, University of California - Santa Cruz, CA 95060 USA
slim@fuzzy.ucsc.edu

[†] LORIA/CNRS, Nancy, France
Yves.Laprie@loria.fr

ABSTRACT

This paper presents a study of the articulatory properties of French vowels using an acoustic-to-articulatory inversion method. The advantage of such an approach is that all the possible articulatory configurations can be studied independently of any articulatory preferences linked with a given speaker. Furthermore, it bypasses the issue of acquiring a vast amount of articulatory data by medical imaging techniques. The inversion method exploits an articulatory codebook, the acoustic precision of which is constant whatever the articulatory region considered. Since the inversion is performed from the first three formants of vowels to recover the seven parameters of Maeda's model the null space of the articulatory to acoustic mapping is explored to recover all the possible articulatory shapes. Applied to French vowels this method allows the different places of articulation to be determined.

1 INTRODUCTION

It is known that speech production exploits compensatory effects (1). In fact, the vocal tract configuration necessary for the production of a vowel can be realized by many different combinations of individual articulators. Thus, an invariant acoustic target can be reached by coordinating articulators to compensate for each other.

Beside the classical way of describing the vocal tract by specifying the position of articulators, it is possible to use geometrical measures. In fact, Stevens and House (2) pointed out that the most important characteristics of vowels from an acoustic phonetic point of view were the position of the main constriction between the tongue and the vocal tract wall and the degree of constriction (cross-sectional area) at that position. Studying and classifying the vowels according to the main constriction can be done by observing real data (MRI and x-ray) and performing a quantitative study as it was done by Wood (3). Wood used X-ray images from 38 speakers to study vowels and found out that there were four possible constriction locations (low pharynx, high pharynx, soft palate and hard palate).

While using real data is the best way to study vowels through the determination of the main constriction, this approach suffers from several limitations. Even though Wood's work was considered the most complete study of the place of articulation for vowels using real data, the data were still limited and thus prevented all articulatory configurations to be covered.

Therefore, Boë et al. (4) used an articulatory model (Maeda's articulatory model (5)) to study the place of articulation for French vowels, which was an important step to study all the vocal tract configurations for a given vowel. However, their work still did not provide a complete set of solutions, since they did not use any inversion method but a limited number of random articulatory configurations. Indeed, it should be noted that the 30000 articulatory configurations used by Boë et al. corresponded approximately to the choice of only 4 samples for each of the articulatory parameters.

A solution to study vowels with respect to the place of articulation is thus to derive articulatory information from speech directly. The advantage is to enable a large amount of vowel examples to be analyzed. However, as mentioned above, it requires an inversion procedure to recover articulatory parameters from speech. Since there is no one-to-one relationship between the articulatory and acoustic spaces (6), most of the inversion methods only provide a partial set of solutions without any knowledge of the inverse solution space coverage. However, The study of articulatory characteristics of vowels requires an inversion method that describes the solution space precisely.

In this paper, we present an inversion method that fulfills this requirement. It is based on an articulatory codebook that adaptatively samples the articulatory space to guarantee a quasi-linear frequency resolution in the acoustic domain. Articulatory entries are structured in the form of a hierarchy of hypercubes. Since the non-linearities of the articulatory-to-acoustic mapping are therefore well represented almost all realistic vocal tract shapes corresponding to an acoustic entry can be recovered.

In the next sections, we present the hypercube codebook generation method and the inversion procedure providing

a complete set of solutions. Then, we present the results of the inversion performed on six French vowels and compared with those obtained by Wood.

2 Inversion method using hypercube codebook

2.1 Articulatory model adaptation

As many others, our acoustic-to-articulatory inversion method exploits the analysis by synthesis paradigm. We used Maeda's articulatory model (5) which represents the vocal tract geometry with seven articulatory parameters: three parameters for the tongue (tongue dorsum, tongue body and tongue tip), two parameters for lips (opening and protrusion), one parameter for the jaw and one for the larynx. It needs to be adapted before being applied to a particular speaker. We used the method proposed by Galvan (7) because it only requires the measurement of formants for a very limited set of vowels. This method computes two scale factors, one for the pharynx, and one for the mouth. Applied to our subject PMM this adaptation gives 1.039 for the mouth coefficient, and 1.038 for the pharynx.

2.2 Generation of the hypercube articulatory codebook

The difficulty of generating an articulatory codebook lies in the presence of non-linearities in the articulatory-to-acoustic mapping: in some articulatory regions, a small variation of one or several articulatory parameters gives rise to large variations of the formants. Conversely, in some regions, large parameter variations do not produce any significant acoustic changes.

Our approach aims at sampling the articulatory space densely only in regions where the mapping is highly non-linear. The results of this adaptative sampling are stored in a hierarchical structure: the leaves correspond to articulatory hypercubes where the mapping is sufficiently linear. The linearity test is evaluated at the center of all segments joining any two vertices of the hypercube considered. The test consists in comparing the formants given by the articulatory synthesizer to those linearly interpolated from formants of the two segment extremities. If the difference between formants synthesized and those interpolated is less than a predefined threshold, the hypercube is considered linear with respect to the articulatory-to-acoustic mapping, otherwise this hypercube is decomposed into sub-hypercubes and the linearity test is repeated for each of these new hypercubes. This procedure is repeated recursively until the linearity is ensured or the hypercube size becomes smaller than a predefined value. The result of these successive decompositions is a hierarchical structure composed of hypercubes of different sizes. The strong point of this codebook construction method is that the complete acoustic behavior of the articulatory model is taken in account. Therefore, the codebook presents the most complete representation of the articulatory space.

2.3 Inversion method and exploration of the solution space

Our inversion method uses the codebook by recovering the possible articulatory vectors for each 3-tuple of formants (the first three formants) of the speech to be inverted. This means that for each acoustic entry, all the hypercubes whose the acoustic image contains the acoustic entry are considered.

We now describe how inverse solutions are found for one of these hypercubes denoted H_c. Let F be the 3-tuple of formants to be inverted. Let P be an articulatory vector (i.e. the seven parameters of Maeda's articulatory model) that is the unknown. The following equation has to be solved: $F - F_0 = \nabla F.(P - P_0)$ where ∇F is the gradient of F calculated at P_0 and F_0 is the formant 3-tuple at P_0.

The general solution of this equation is given by a particular solution plus any vector from the null space. This means that adding a linear combination of the basis vectors of the null space does not change formants. The SVD (*singular value decomposition*) method as described in (8) gives one particular solution and an orthonormal basis of the null space.

In our case, as $M = 3$ (3 formants) and $N = 7$ (7 articulatory parameters), the null space dimension is generally 4. To retrieve all the solutions, the null space has to be sampled (further details can be found in (9)). The number of inverse solutions and their accuracy depend on the sampling of the null space.

3 Experiments

In Tab.1, we present the first three formant values used for the 6 French vowels. These values are obtained by extraction of the formants from a speech corpus of isolated vowels uttered by our subject PMM for whom we adapted Maeda's model.

Vowels	$F1$	$F2$	$F3$
a	601	1526	2266
e	375	1860	2612
i	317	1857	2675
ɔ	418	1225	2125
u	390	939	2115
y	257	1818	2253

Table 1: Three first formant values (Hz) for six French vowels used for the inversion experiments.

We applied the inversion procedure on the acoustic data and obtained several vocal tract shapes for each vowel. To check the accuracy of the inversion results, we resynthesized the articulatory data and we compared the measured acoustic data to those synthesized from the articulatory data. As one can notice in Tab. 2, the acoustic preci-

sion is excellent and a large number of vocal tract shapes is found.

We now present the results with respect to three parameters: cross-sectional area of the main constriction (A_c, cm^2) also called degree of constriction, position of the main constriction in the vocal tract (X_c, cm) also called place of articulation and the cross-sectional area at the lips (A_l, cm^2) also called lip opening. These parameters are obtained by retrieving the vocal tract section where the cross-sectional area is minimal. We consider neither the constriction formed at the lower part of the pharynx (close to the larynx at 2 cm from the glottis) nor that formed by lips.

Vowels	$\triangle F1$	$\triangle F2$	$\triangle F3$	# of solutions
a	5	5	0	134211
e	6	7	7	156541
i	10	14	7	64724
ɔ	8	5	3	90140
u	24	6	2	12254
y	7	9	13	8595

Table 2: The result of the resynthesis of the inversion solutions. Where ($\triangle F1$, $\triangle F2$, $\triangle F3$) are the errors between the measured and resynthesized formants. The last column gives the number of inversion solutions obtained for each vowel.

For each vowel, the results are presented in three different forms (see Fig. 1): lip area with respect to the main constriction area, lip area with respect to the main constriction position, and constriction area with respect to the position of the main constriction. The position of the main constriction varies between 0 (glottis) and 17.5 cm (lips).

We present the results for the vowel /a/ in Fig. 1. The main constriction is situated in different regions: pharynx, upper pharynx, soft and hard palate. The solutions of the first and the third classes (the constriction lies within the pharynx or the palate) form 94% of all the solutions. The constriction area for the first class varies between 0.23 and 4 cm^2. For the second class (composed of 5% of all the solutions), constriction areas are very high (greater than 3 cm^2). The third class results from a presentation artifact and actually corresponds to places of articulation very close to lips. For the three true classes, the mouth opening varies widely but the area at the lips is generally bigger than 2 cm^2. Considering that reasonable constriction areas should not be greater than 3 cm^2 one can note that the inversion procedure finds a large number of inverse solutions that are probably not realistic even if they are generated by the articulatory model. Together with constraints proposed by Boë et al. (4) this provides another source for adding constraints to restrict the range of Maeda's articulatory parameters.

For all the experiments we carried out on vowels, we notice that variability of the vocal tract shapes is smaller than the variability of the articulatory parameters. In fact, the main constriction places are located in 3 different regions for the

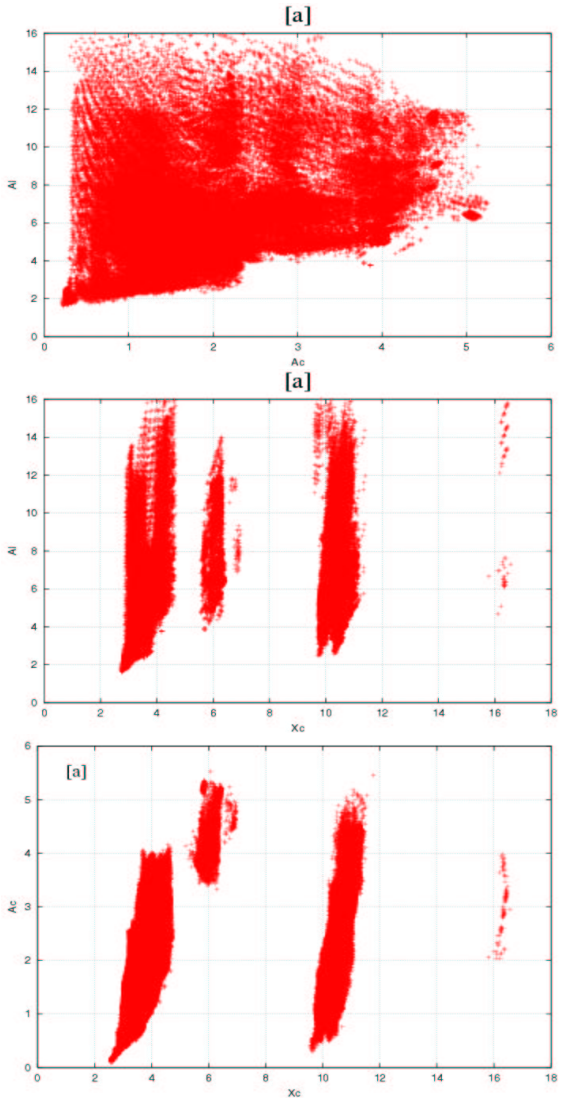

Figure 1: Representation of the vowel /a/ in the planes X_c/A_c, X_c/A_l and A_c/A_l.

vowel /a/ (by discarding the smallest set of solutions whose constriction almost corresponds to lips), three for the vowels /ɔ/ and /u/ and "almost" one region for the vowel /e/, /i/ and /y/. This confirms that there exist very precise places of articulation for vowels that are imposed by acoustic constraints.

4 Conclusion

Compared to constriction areas estimated from Xray images by Wood it appears that areas obtained by inversion range over a larger interval. This means that the inversion procedure recovers extreme articulatory configurations that should be discarded. Considering that the highest constriction area observed by Wood for /a/ is close to $2cm^2$ we could restrict the set of reasonable articulatory configurations for /a/ to the part of Fig. 1 X_c/A_c below 2.5 cm^2. Thus this gives only two places of articulation for /a/.

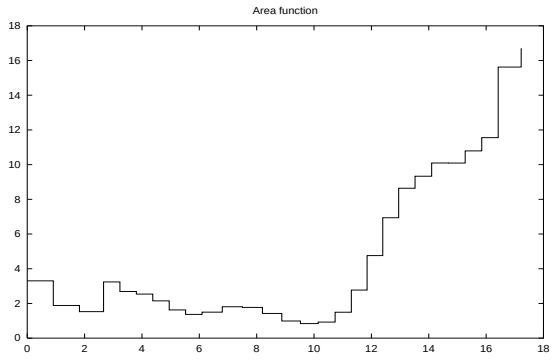

Figure 2: Example of an extreme area function at lips for /a/

The spreading of the constriction areas together with the very high area function obtained for some articulatory configurations (see Fig. 2 at lips for instance) correspond to area function regions for which the formant sensitivity vanishes. Thus, increasing the area does not change formants, and the only limit is given by the range of the corresponding articulatory parameters.

In this paper we adapted Maeda's model to a given speaker. The places of articulation found by inversion cannot represent the speaker variability that could be observed if several speakers had been taken into account. Indeed, it can be noted that second and third formants of /a/ are rather low for this speaker.

Furthermore, the articulatory model itself has been obtained from X-ray images of a given speaker. Thus it can involve some specific articulatory characteristics of the speaker used to build the model.

However, results obtained correlate well with those found by Wood. As we want to obtain results as independent as possible from any speaker we will generalize these results to generic male and female speakers obtained by varying mouth and pharynx scale factors. This will enable us to analyze how places of articulation are related to different anatomical configurations.

References

[1] B. Lindblom, J. Lubker, and T. Gay, "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation," *J. Phonetics*, vol. 7, pp. 147–161, 1979.

[2] K. N. Stevens and A. S. House, "Development of a quantitative description of vowel articulation," *Journal of the Acoustical Society of America*, vol. 27, pp. 484–493, 1955.

[3] S. Wood, "A radiographic analysis of constriction locations for vowels," *Journal of Phonetics*, vol. 7, pp. 25–43, 1979.

[4] L.-J. Boë, P. Perrier, and G. Bailly, "The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion," *Journal of Phonetics*, vol. 20, pp. 27–38, 1992.

[5] S. Maeda, "Un modèle articulatoire de la langue avec des composantes linéaires," in *Actes 10èmes Journées d'Etude sur la Parole*, Grenoble, Mai 1979, pp. 152–162.

[6] B. S. Atal, J. J. Chang, M. V. Mathews, and J. W. Tukey, "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique," *Journal of Acoustical Society of America*, vol. 63, no. 5, pp. 1535–1555, May 1978.

[7] A. Galván-Rdz, *Études dans le cadre de l'inversion acoustico-articulatoire : Amélioration d'un modèle articulatoire, normalisation du locuteur et récupération du lieu de constriction des occlusives*, Thèse de l'Institut National Polytechnique de Grenoble, 1997.

[8] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1989.

[9] S. Ouni, *Modélisation de l'espace articulatoire par un codebook hypercubique pour l'inversion acoustico-articulatoire*, Thèse de l'université Henri Poincaré, Nancy, Dec. 2001.