

COARTICULATION AND DEGREES OF FREEDOM IN THE ELABORATION OF A NEW ARTICULATORY PLANT: GENTIANE

Vilain Anne, Abry Christian, Badin Pierre

Institut de la Communication Parlée (ICP), France

e-mail: {vilain, abry, badin}@icp.inpg.fr

ABSTRACT

A new articulatory model GENTIANE — elaborated from an X-ray film built on a corpus of VCV sequences performed by a skilled French speaker — enabled us to analyse *coarticulation* of main consonant types in vowel contexts from a *degrees of freedom approach*. The data displayed an overall *coarticulatory versatility*, except for an *absolute invariance* in the labio-dental constriction point. For consonant types recruiting the tongue, the variance explained by the degrees of freedom of the model evidenced specific *compensation strategies*: tongue tip compensation betokened the common coronal status of the dental plosive and the post-alveolar fricative; whereas tongue dorsum compensation signed the dorsal nature of the velar plosive.

1. INTRODUCTION

We have elaborated a new anthropomorphic articulatory model with degrees of freedom in excess, GENTIANE, designed to be integrated in a plant-control robotic framework. This linear articulatory model of speech production was derived from an X-ray film, through statistical analysis [6, 7].

Such a plant modelling offers a *degrees of freedom approach in the study of coarticulation*. It explains the generation of articulatory final or end products, say *distal* configurations (e.g. lip constriction), by a linear combination of articulator degrees of freedom (e.g. mandible and lips), say *proximal* effectors. Hence the coarticulation issue can be formulated as a threefold question. First *distally*: given a set of observed contours, what regions of interest for coarticulatory variability are explained or not by the model-generated contours? Second *proximally*: what are the patterns of variance in control parameters for different coarticulatory types, i.e. which commands are frozen and which are free? Third, in the *proximal-to-distal relation*, what control parameters interact to save stability in a given distal region throughout coarticulation? This last question addresses the *equifinality* issue. Ultimately it will lead to ask why stability is preserved or not, for articulatory-acoustic purposes.

In this study, since vowel targets appeared as a whole to be rather stable across consonantal contexts, priority was given to the modelling of consonant variability and stability in vowel coarticulation.

2. THE PLANT: GENTIANE

2.1. Articulatory Database

Our study is based on a 50 images/second X-ray high quality movie of a French speaker in profile, synchronised with a front-view video film and audio tape. Four lead pellets, on the tongue dorsum, blade and tip, and on the chin of the speaker, helped us detect more precisely the blurred contours of the surface of the tongue, as well as the sub-lingual cavity.

The corpus performed by the subject consisted in Vowel-Consonant-Vowel sequences, with consonants [b, d, g, v, Z] ([Z] stands simply for the IPA voiced post-alveolar fricative), and the cardinal vowels [i, u, a, y]. It was designed so as to compose the most representative “phonetic score”, in order to allow a sufficient exploration of the degrees of freedom of the vocal-tract within only 30 seconds of exposure.

The superposition of all organs in an X-ray picture prevents a reliable automatic detection of contours. Therefore contours were manually traced for the whole vocal-tract, and then digitised, which permitted a large amount of automatic measurements. A semi-polar grid was superimposed on each of the 1282 tracings now available, yielding a vector of 27 vocal-tract sagittal sections per vectorised image (from the larynx to the tongue tip). In addition to the basic Jaw Height coordinate between the two incisive teeth, other measurements were performed, essentially to capture lip configurations.

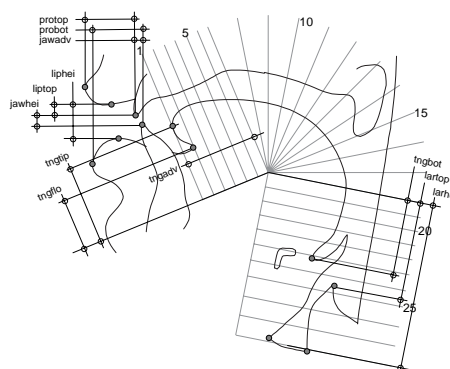


Fig. 1: Grid and afferent articulatory measurements.

2.2. Elaboration of the Plant

From this database, we extracted the degrees of freedom of our model, following a statistical method inspired from Maeda's analysis [4]. An articulatorily-driven Overall Principal Component Analysis (PCA) was run on the data. (i) The effect of Jaw Height (JH) was subtracted from the whole variance of the contours around the neutral position. (ii) PCA was then run on the body of the tongue, without taking into account the apex, so as to extract two components that correspond to Tongue Body (TB) and Tongue Dorsum (TD) movements. (iii) PCA run on the apex yielded the Tongue Tip (TT) component. (iv) Finally Tongue Advancement (TA), Larynx Height (LY), Lip Height (LH), Lip Protrusion (LP), and upper Lip Vertical elevation (LV), were all kept as command parameters. As a result, 9 degrees of freedom were obtained as control parameters for the plant, coherently with another ICP model BERGAME [2].

In addition an inversion procedure was developed taking advantage of the lead pellets, in order to use this articulatory model for the reconstruction of contours of the whole vocal-tract, with a few points array tracked by electro-magnetic articulography, thus offering the opportunity to extend the database on the same subject without further exposure to X-rays [1].

3. DEGREES OF FREEDOM IN THE COARTICULATION OF CONSONANTS

For each consonant type, we examined the acoustically central phase in as many vowel contexts as available in the corpus.

3.1. Method

We based our analysis on three different types of complementary documents (shown below, when needed).

Sagittal contours (articulatory *distal* end-products). Configurations of the same consonant type were superimposed for different vowel contexts (e.g. **Fig. 2**). These figures are actual vectorised tracings from the initial corpus. They give an overview of the absolute variability of the configurations, without an analysis of the contribution of each articulator to these final settings. To obtain this contribution they have to be compared with model-generated contours.

Command parameters panels (*proximal* effectors). Each command is normalised between ± 3 standard deviations. Within the coarticulatory contexts under examination, such a display gives the standard deviation range around the mean for each parameter, in the following order: JH, TB, TD, TT, TA, LY, LP, LH, LV (see. **Fig. 4**). We insist on the fact that these panels show command parameters, i.e. normalised values for the action of each degree of freedom, not an effective distance in the configuration to be produced. This normalisation enlightens the actions of articulator commands, whatever the resulting spatial displacement of

end-effectors. E.g. we know that a slight movement of the tongue tip may be as crucial for the accuracy of the articulation as a large movement in tongue body. So the corresponding commands preserve the relative efficiency of all articulatory movements within the speech working space.

Vocal-Tract variance explained by command (*proximal*) parameters along the (*distal*) configurations of the vocal-tract (grid section numbers). Here we look for the contribution of each control parameter in the generation of the different coarticulated configurations of the consonant type: in other words the geometrical variance explained by each articulator command. "Vocal-tract variance figures" show the non-normalised standard deviation of the data (in cm), on every line of the grid. It is thus possible to visualise the local action of a parameter in a region of the vocal-tract, and the interactions between articulator commands for the achievement of the desired consonant setting in this region.

On these "VT variance figures" (e.g. **Fig. 5**), the first bold line is the total standard deviation of the sub-corpus contours around the neutral configuration. Then we have the variance left after having subtracted the effect of each one of the first 5 command parameters, in the following order: JH, TB, TD, TT, and LY, being the lower bold line. (The remaining parameters were not taken into account here, since TA would not have any visible effect on this mid-sagittal deviation, and the lips were measured outside the grid).

3.2. Labial Constrictions

"Contour figures" **2a** & **2b** show that [b] and [v] impose only one point to be constrained, namely the labial constriction. As long as this condition is fulfilled, the other articulators are quite free to achieve the vowel gesture. This is the most obvious for the tongue. But both consonant types allow relatively large variations in Larynx Height in order to produce rounded vowels in synergy with upper and/or lower lip protrusion: the lowering of the larynx is acoustically relevant, since it adds to the lengthening of the vocal-tract.

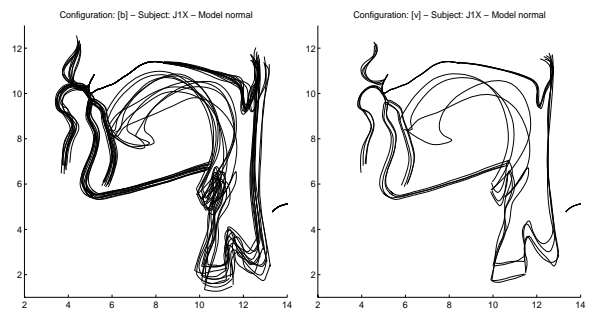


Fig. 2: Sagittal contours in [b] (**2a**) & [v] (**2b**) in [i, a, u, y] contexts (cm).

[b] allows some variability in lip protrusion in order to keep on the vocalic protrusion of [u] or [y]. The exact place of closure does not seem to matter as much as in [v], for which an absolute precision was observed in the labio-dental constriction. However this does not seem to be actually motivated by an articulatory-acoustic constraint, since we

know that a slightly modified place of articulation would not alter basically the quality of the fricative. Besides, such an absolute invariance in the lower lip was not found for the other speakers which were recorded and modelled in the same vein. While lip elevation is obviously needed for [v] production, whether *protrusion* of the two lips has to be yoked or not is still a modelling issue.

3.3. Coronal Constrictions

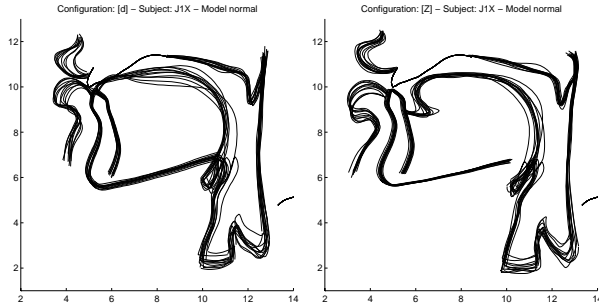


Fig. 3: Sagittal contours of [d] (3a) and [Z] (3b) in [i, a, u, y] contexts.

As concerns lip actions, [d] seems at least as free in coarticulation as [g] (below). The protruded fricative [Z] is less variant specifically in the Lip Protrusion degree of freedom (see “command panel” in **Fig. 4b**), though more than [v] anyway.

In contrast with [b] and [v], [d] and [Z] are both consonant types where the tongue is recruited for the constriction. While being both coronal types, they differ in the degree by which Tongue Tip interacts with Jaw Height and Tongue Body. In fact these two consonant types show evidence of compensation between articulators, in order to fulfil equifinality for their specific articulatory-acoustic constraints.

Figs 5a & 5b clearly show this strategy. For [d], about grid-section n°26, the total variance is small, because of the constraint on place of articulation. Jaw Height can be seen increasing the variability at that point: standard deviation after subtraction rises above the total variance. This could “endanger” the achievement of the closure. But Tongue Tip compensates for this increase and brings the variance down. It is actually the coarticulatory effect of the vocalic configuration that moves more or less Tongue Body to and fro into the mouth, and Tongue Tip has to keep its own upward movement to make up for that motion.

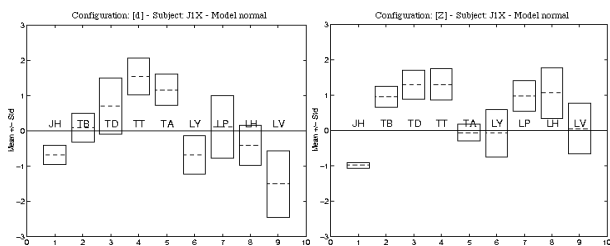


Fig. 4: Command parameters panels for [d] (4a) & [Z] (4b).

For [Z], the overall configuration is more constrained, far back from the apex into tongue dorsum. [Z] has a remarkably stable Jaw Height (**Fig. 4b**): it shows the smallest standard deviation of all the observed consonants. This degree of freedom is practically frozen, which does not impede a certain synergetic *protrusion* of the jaw. This forward translation is visible on the contours in **3b** (but lost in the next figures, since the model has been elaborated with only one degree of freedom for the jaw, and this has to be kept in mind throughout the observations).

Anyway, Jaw Height does not add variance at the apex in [Z], but Tongue Body does (**Fig. 5b**, about grid-section n°25). And the same Tongue Tip reaction as in [d] appears. In short: though the perturbation does not come from the same articulator, the same compensation operates.

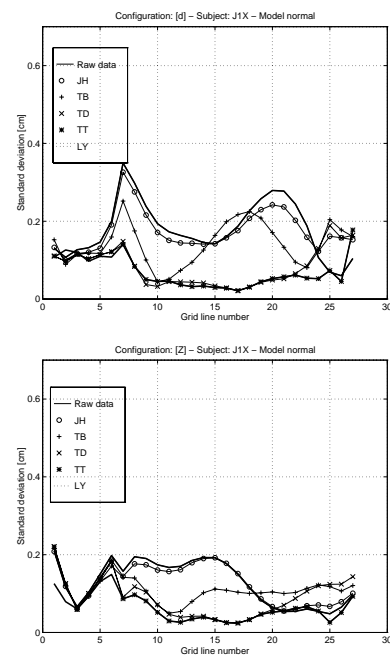


Fig. 5: VT variance explained by command parameters for [d] (5a) & [Z] (5b). (NB: grid origin was set at the larynx).

3.4. Dorsal Constriction

As compared to other consonant types seen previously, [g] shows a global variability of the whole vocal tract (**Fig. 6a**). More than [b], this type is versatile in the sense that it can afford a fair range in the location of its closure: points of contact differ by 2 cm between the palatal [i] and velar [u] contexts. The mass of the tongue being recruited for the vowel shaping, a dorsal constriction location is quite difficult to control independently. The resulting coarticulation strategy is that the tongue body is somewhat locked to vocalic locations, and the tongue dorsum acts to achieve the dorsal constriction anyway.

This is to be observed on **6b**: variability in the vocal-tract is almost exclusively explained by TB, except about grid-section n°19, where it is TD in turn that explains most of the variance. Though the same perturbator as in [d] is acting, i.e.

Jaw Height, a different compensatory end-effector operates. Whereas Tongue Tip reaction evidenced the coronal status of [d], Tongue Dorsum compensation signed the dorsal nature of [g].

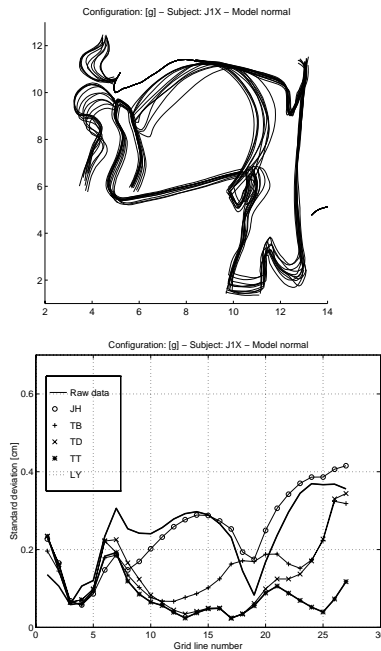


Fig. 6: Sagittal contours for [g] in [i, a, u, y] contexts (6a); VT Variance explained by command parameters (6b).

4. CONCLUSION

Articulatory modelling of a French-speaking X-ray database enabled us to analyse coarticulation of main consonant types in vowel contexts through a degrees of freedom approach. These *proximal* commands were used to explain the variance of the final, *distal*, vocal-tract configurations. Using these *proximal-to-distal relations* we evidenced compensatory strategies meant to achieve specific equifinality for the different consonant types under examination.

Apart from the versatility of the labial plosive, which just needs to maintain its constricted state, be it protruded or not, constraints on *place* category (exceptionally invariant for the labio-dental in this subject) are essentially achieved by *compensatory* manoeuvres intrinsic to each consonant type. Thus when the dental plosive is perturbed by *jaw* variation, the *tongue tip* compensates; and the same tip compensation happens when the post-alveolar fricative is perturbed by *tongue body* action. In contrast with this typical *coronal* behaviour, the *dorsal* status of the otherwise versatile velar or palatal plosive is guaranteed by *tongue dorsum* compensation when perturbed by the *jaw*.

Such an approach allows quantitative comparisons with other models, elaborated on a set of subjects, with a similar methodology [3].

This is a first step towards biomechanical modelling of coarticulation for consonants as well as for vowels [5].

5. ACKNOWLEDGEMENTS

The X-ray film was realised by the Schiltigheim Hospital staff through a collaboration with the Phonetic Institute in Strasbourg. We thank: Gilbert Brock for technical assistance, Péla Simon and Jean-Pierre Zerling for advisory support, and Tahar Lallouache (formerly at ICP) for pellet placing and video.

6. REFERENCES

1. Badin, P., Baricchi, E., Vilain, A., "Determining tongue articulation: from discrete fleshpoints to continuous shadow", *Eurospeech*, Rhodes, Greece, 1, pp. 47-50, 1997.
2. Badin, P., Gabioud, B., Beautemps, D., Lallouache, T.M., Bailly, G, Maeda, S., Zerling, J.P., Brock, G., "Cineradiography of VCV sequences: articulatory-acoustic data for a speech production model", *15th ICA*, Trondheim, Norway, pp. 349-352, 1995.
3. Bailly, G., Badin, P., Vilain, A., "Contribution de la mâchoire à la géométrie de la langue dans les modèles articulatoires statistiques", *XXIIèmes Journées d'Études sur la Parole*, Martigny, Switzerland, pp. 287-290, 1998.
4. Maeda, S., "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model". In Hardcastle, W.J. and Marchal, A. (Eds), *Speech Production and Modelling*, pp. 131-149. Kluwer: Acad. Publishers, 1990.
5. Sanguineti, V., Laboissière, R., Ostry, D.J., "A dynamic biomechanical model for neural control of speech production", *J. Acoust. Soc. Am.* 103 (3), pp. 1615-1627, March 1998.
6. Vilain, A, *Un nouveau modèle articulatoire pour la synthèse et le contrôle robotique de la parole: Gentiane*. DEA Sciences du Langage, ICP, Grenoble, 1997.
7. Vilain, A., Abry, C., Badin, P., "A propos des degrés de liberté dans la coarticulation d'un locuteur français filmé 50 fois par seconde", *XXIIèmes Journées d'Études sur la Parole*, Martigny, Switzerland, pp. 311-314, 1998.