ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

MOTOR EQUIVALENCE EVIDENCED BY ARTICULATORY MODELLING

Vilain Anne, Abry Christian, Badin Pierre

Institut de la Communication Parlée (ICP)˝- CNRS UPRESA 5009
Université Stendhal, BP 25, Grenoble cedex 9, France
avilain@icp.inpg.fr
http://www.icp.inpg.fr

10.21437/Eurospeech.1999-46

ABSTRACT

We present a method for the analysis of motor equivalence on two French corpuses, evidenced by articulatory modelling. This processing enables us to make out the individual actions of each degree of freedom of the vocal tract. We intend to define the phonetic types we are studying with a combination of degrees of freedom recruited. Such a characterisation makes it possible to account for coarticulatory processes in a coherent way.

This method brings to light compensatory strategies, striking strategies of preservation of vocalic configurations in some consonants by one speaker only, and a large divergence in the strategies used by the two speakers.

Keywords: speech production, coarticulation, articulatory modelling, motor equivalence

1. INTRODUCTION

What is the use of articulatory modelling in the study of motor equivalence ? Although coarticulatory effects and the resulting variations in displacements of jaw or tongue can be clearly observed on sagittal views of the vocal tract alone, this global observation does not allow the making out of the different contributions of each individual articulator : particularly, in a zone where stability is needed for a constriction, a global contour does not yield any information about the combinations that are implemented to make up for perturbations by antagonist gestures, in other words, to reach equifinality for all realisations of a given sound in any coarticulatory context.

We have used two global linear articulatory models based on two cineradiographic corpuses of two French speakers: this subject-oriented modelling allows to extract the degrees of freedom of the articulators, while preserving the conformation and individual synergetic strategies of the speaker.

Speech production is therefore studied under the light of the different actions of each parameter, which is particularly meaningful for the tongue: its global contour alone is hardly interpretable in terms of different articulator contributions: apex, dorsum, body.

The methodology we have elaborated allows to decompose a vocal tract configuration into the individual contributions of each articulator. We carry on here the study presented in Vilain [6] with a more detailed analysis: coarticulatory effects, variability and strategies compensating for natural perturbations, that is in non-artificially perturbed speech, are explored, in consonants, in order to establish a coarticulation typology and to propose explanations for the different behaviours observed. We finally question such issues as: (i) how far does speech production consist in a consonant gesture superimposed on a continuous vowel-to-vowel gesture (Öhman [4]) ?; (ii) what is the acoustical pertinence of compensatory manoeuvres of the lingual-mandibular complex: are they implemented in order to limit the variations in vowel-consonant or consonant-vowel formant transitions that would otherwise be induced by too great a coarticulation, therefore to reduce acoustic variation (Edwards [1]) ?

2. DATA AND MODEL

The data consist in two cineradiographic corpuses of two French speakers uttering VCV sequences with C=[b,d,g,J,v], and V=[i,y,u,a]. From these data, two linear anthropomorphic articulatory models have been elaborated.˝ The degrees of freedom of the speaker's vocal tract are emerged with an articulatorily-driven Principal Component˝ statistical Analysis and used as command parameters of the model.

These nine parameters, namely Jaw Height,˝ Tongue Body, Tongue Dorsum, Tongue Tip, Lip Height, Lip Protrusion, Lip Vertical Elevation, Larynx Height, and Tongue Advancement, can be assumed to represent fairly well the degrees of freedom of the articulators of the vocal tract.

The data and model have been described extensively in Vilain [6].

3. ELABORATION OF A METHOD

Maeda [3]˝ suggested in 1990 that "the analysis of the conventional cineradiographic data using an articulatory model can provide a deeper insight into the speech production process, and […] it can help us to˝ understand better the inherent characteristics of speech, the variability, that makes speech so interesting to study". We˝ have followed and developed this view, in using the model to extract combinations of articulators for different phonetic types.

Different˝ ways of visualising the recruiting of the degrees of freedom of the vocal tract were investigated. We envisaged several kinds of graphs, each of them yielding a different type of information, whether they described final results, i.e. global articulatory configurations, or bricks in˝this final product, i.e. individual contributions of each degree of freedom to the˝whole configuration.

We have chosen to study first the occurrences of consonants in a symmetrical vocalic context, to facilitate the association of gestures either to the vowel or the consonant. But actually, the corpuses are not exactly similar, therefore the P1X corpus lacks some information, such as˝ the symmetrical˝ [u-u] and [y-y] contexts.

3.1. Comparison of sagittal contours

We started with global contours of the vocal tract, thus studying final articulatory configurations. A superposition of sagittal contours allows to visualise the˝points of minimum and maximum variation, but not to attribute this fact to any one of the different articulators. It can be used as a guideline for the next studies, but is not meaningful enough in itself (Vilain [5]).

3.2. Global observation of contributions

To explore the strategies of coarticulation, we had to have an more inner view of the actions of the different degrees of the tongue. We first tried to gather in a synthetic view the global behaviours of degrees of freedom for one phonetic type in different contexts, to have a hint at what kind of coarticulatory processes were characterising this very type. This observation is exemplified on figure1.

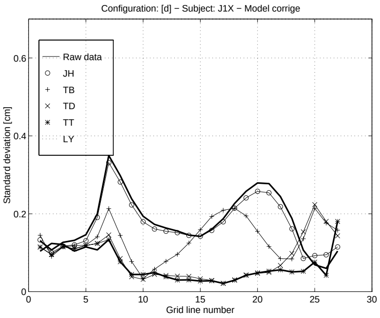

Fig. 1: Vocal Tract variance explained by command parameters for [d] (NB: the grid origin is at the larynx).

On this figure, we showed the geometrical vocal tract variance explained by each of the command (*proximal*) parameters along the (*distal*) configurations of the vocal-tract (grid section numbers). On these “VT variance figures”, 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.

This view tells where on the grid-lines the minimums of variance are found, and to which of the command parameters it is mostly attributable. For example, on figure1, the minimum is logically in the coronal region. What is interesting is that Jaw Height can be seen increasing the variability at that point: standard deviation after subtraction rises above the total variance. And 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.

Although such an evidence for a compensatory strategy is valuable, it appeared to be insufficient, since the behaviours are very much differentiated according to the neighbouring context. The synthetic figures give then a direction to be explored but not a precise view of the actions actually produced.

3.3. Contributions for single sequences

The computation of the individual actions of the first four parameters can be drawn for one item at a time. It can be used to observe the unfolding of actions during the transition from one item to another, particularly from a vowel to a consonant in a VCV sequence. The information extracted from this view are: which degree of freedom is recruited for the vowel, which one for the consonant, and how is the transition carried out ?

The next step was therefore to observe separately each sequence of one form in a precise phonetic context. This allowed to view precisely what the synergetic strategy was in each case. The sagittal contour is, in a way, decomposed into the individual contributions of each of the first four parameters. The point is to define the goal of motor equivalence, not as a geometric final configuration, but as a needed combination of degrees of freedom.

Undeniably such a study of coarticulation is at the risk of the model. We have to hypothesise, first of all, the adequacy of the model to reproduce the actual behaviour of degrees of freedom. We think such a postulate is the only way to represent otherwise invisible mechanisms.

4. CONSONANTS

1. Bilabials

Bilabial consonants, especially [b], can be thought to be unspecified for lingual gestures, therefore not to impose any change on the vowel-to-vowel gesture but for the lip closing, maybe associated with a jaw raising meant to make the lip closure possible.

This is found in most realisations of [b]. Therefore the variability of other non-specified articulators is expectedly large. The consonant is largely coarticulated with its vocalic context. This is obvious if we look at the superimposed contours of all realisations of [b]: The constraint is the lip closure, which itself has some variability: it is not a very fixed position of the lips and it can vary with, for instance, the protrusion of the vowel.

So the whole contours show a large variability. Now a more precise study of each utterance renders a number of intriguing facts.

For [y], the jaw raises lightly, TB remains in its vocalic position, but the apex lowers through the added actions of TD and TT.

[i] shows the same pattern, with a raising of the jaw, a diminishing but still positive action of TB, and a counteraction of TD and TT to bring the apex lower. This lowering movement could be seen as a manoeuvre to avoid a possible occlusion, yet the distance between the apex and the palate for [b] in [ibi] is the same as in [y] before [yby], it cannot therefore be considered as a critical distance and the lowering must have another significance, and may result from an acoustic constraint.

[ubu] shows no particular change from the vowel to the consonant, except for a slight raising of the jaw.

Now, figure 2 shows an [aba] sequence as articulated by J1X. The samples are extracted at the centre of the consonant and vowels. The first figure attests the typical [a] articulation, with the jaw open only. The third one is very similar to the first one. The middle one shows the contributions in [b], and there stands an interesting point: as we said earlier, a bilabial consonant would not be supposed to be specified for the action of the tongue, and therefore its production in a context that does not recruit the tongue either would be thought to be carried out with the jaw and lips only. But strikingly, [b] appears here to necessitate a complete recombination of the articulators, aimed at recomposing the open shape of [a]: the very low jaw height necessary for the production of [a] is brought back to zero for the consonant; yet the body of the tongue is not passively raised by this movement, as could be supposed. Instead we observe a reorganisation of the articulators, whose combined actions reconstruct the vocal tract shape of [a], without the contribution of the jaw. Rather unexpectedly the new position of the tongue compensates for the high position of the jaw implied by the lip occlusion. And this new combination cannot be considered as an anticipation towards a new [a] configuration, since the second [a] is produced exactly as the first one was.

The recombination is therefore specific for the consonant. This counteracts the idea of a passive V-V gesture on which the consonant is superimposed, since there exists an active manoeuvre to keep the vowel stable. What is the ground for such a constraint ? Could it be acoustic ? The study of the acoustic corpus is necessary to propose an explanation.

This observation joins Engstrand's finding [2], that the vertical movement of the tongue in /ipi/ shows troughs during consonant, which implies that bilabials might be specified for tongue position, maybe to optimise aerodynamic conditions.

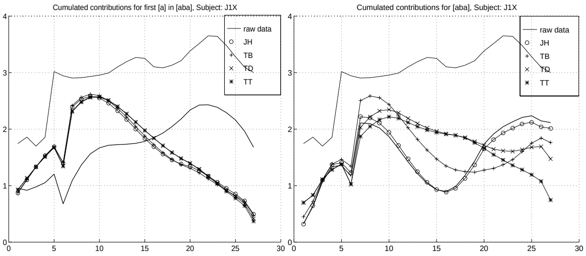

Figure2: [a] and [b] in an [aba] sequence by speaker J1X

This behaviour is found in both labials in this subject, but not that clearly in the second subject. Why ?

The second subject does not show exactly the same pattern: [a] is produced with JH and TB, and this configuration is maintained during the production of the consonant, except that JH is then less open. There is not actually a recombination, but only a change in amplitude (figure3).

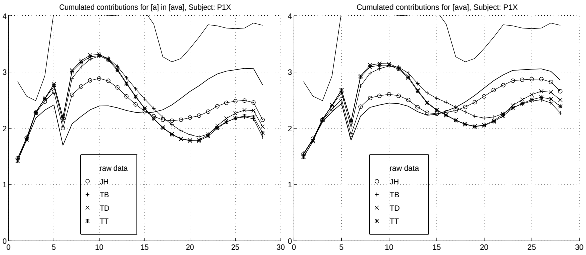

Figure3: [a] and [v] in [ava] sequence by speaker P1X

2. Coronals

For speaker J1X, both coronals in the corpus show a pattern of recruiting of all articulators, but not to the same extent.

[d] implies an action of JH, which reaches the right height to help complete the apical occlusion, and of course an action by TT to perform the occlusion.

But then, different patterns of action of the tongue are used: for [idi] a combination of TB and TD is implemented to slightly lower the tongue at the rear of the occlusion. For [ydy] and [udu] the actions of these two tongue parameters are brought to zero during the consonant, though in [udu] they are largely active in the vowel (figure4). Actually there seems to have no coarticulation at all in this sequence, except for the lips. For [ada], the TB/TD combination required by the open vowel is kept throughout the vowel, while the jaw and the tongue tip are high. The combination is even exaggerated to compensate for the position of the jaw and maintain a really open configuration. As seen for the labials in [a] context, this behaviour is also a quite unexplained one. If the position of TB and TD had remained exactly the same, it would only have meant coarticulation: the vowel is preserved as long as possible and the consonant gesture is superimposed on that. This one is different. We can suppose that the explanation is acoustic and that this additional gesture helps produce clear formant transitions, but this has to be verified, and also confronted with the timing of these gestures. Anyway, why then would this strategy not be applied for the back vowel, whose articulatory configuration is completely lost during the consonant ?

The second corpus, P1X, does not show the pattern of preservation of the open vowel throughout the consonant, but has JH, TB and TD brought back to zero.

For [idi], it seems that the jaw alone performs the slight change in the position of the tongue, from anterior vowel to anterior constriction.

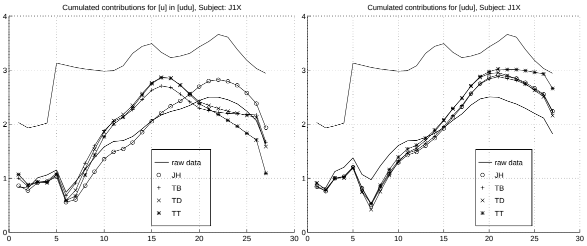

Figure4: [u] and [d] in [udu] sequence by J1X

[J] does not only need an apical constriction as in [d], but also a long laminal "channel", therefore the TB and TD parameters are not only neutralised here, but rather they actively participate to the shaping of the tongue. Their role is to lower the mass of the tongue, to compensate for the very high position of the jaw required by the precision of the constriction. Compensation is observed in all contexts, and used to flatten the tongue.

Now speaker P1X does not show the same strategy. He does seem to need such a flattening of the tongue.

For the production of [J] in [a] context, P1X needs to have a very high JH, then TB brings the tongue down, and then TT reaches the constriction. On the contrary, in [i] context, the laminal constriction is produced almost only with the TB parameter, which is rather striking (figure5). Actually, if we look at the global sagittal views of [J] in different vocalic contexts for both speakers, P1X shows much more variation than J1X for the shape and place of the tongue. Once more, our two speakers show very differentiated behaviours ? The study of the acoustic results should give us a few answers. Anyway, it is interesting to observe that two speakers of the same language use so different strategies for the same phonetic product. Why does J1X impose so many constraints on his production?

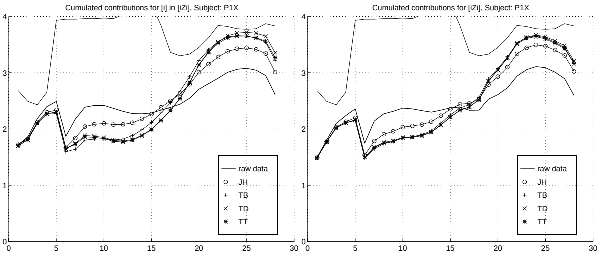

Figure5: [i] and [J] in [iJi] sequence by speaker P1X.

3 Dorsals

For the production of the velar occlusive [g], the only intended equifinality is the dorsal occlusion. Therefore the consonantal activity appears here as a superimposition, on the vocalic gesture, of an upward movement of the dorsum. Such a behaviour leads, logically, to a diversity of the points of occlusion. Here there is no need for any compensation activity, since TD is able to perform the occlusion alone (see figure 6 for [ugu]), without disturbing the positions of the other parameters, except for /a/, of course, where JH is needed to reach the appropriate height for the occlusion.

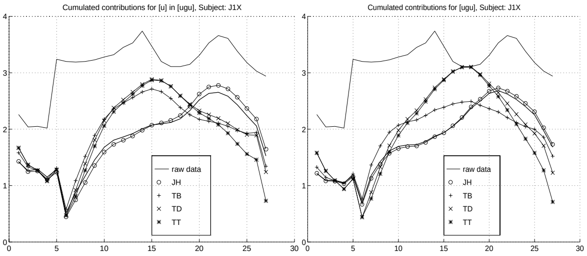

Figure 6: [u] and [g] in a [ugu] sequence by J1X.

P1X uses this strategy for [igi], where TD alone changes the configuration, but rather a combined action of TB and TD to complete the occlusion in [aga].

7. CONCLUSION

The method we have elaborated constitutes a new way of characterising degrees of freedom in the coarticulation of VCV sequences. It brings to light compensatory strategies, not only between lips and jaw, or tongue and jaw, as had already been observed, but also between the different articulators of the tongue.

This observation raised some interesting questions, such as the seemingly superfluous preservation of vocalic configurations in some consonants by speaker J1X only, and the large divergence of strategies used by the two speakers, and all this directs us towards a close analysis of the acoustic variability associated to such different behaviours.

However, one could suppose that the complex strategies implemented by J1X might fade in a hypo-articulated corpus, therefore we intend to record new data with these two speakers. We also need to harmonise the two corpuses.

The next question is now: How should vowels be characterised ? Vowels seem quite static across consonantal contexts and the point is to determine the goal to be reached in their production, whether it is a spatial configuration, a precise place of constriction, i.e. the highest point of the tongue. It is also to be determined whether they imply compensation strategies, for instance between lip protrusion and larynx height for rounded vowels.

8. REFERENCES

[1] Edwards, J. (1985), Contextual effects on lingual-mandibular co-ordination. *Journal of the Acoustical Society of America 78* (6), pp.1944-1948, December 1985

[2] Engstrand, O. (1983), Articulatory co-ordination in selected VCV utterances: a means-end view. *Reports from Uppsala University Department of Linguistics*, 10, 1-145.

[3] Maeda, S. (1990), Compensatory articulation during speech; evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model, *Speech Production and Speech Modelling,* Eds. Hardcastle and Marchal, pp131-150.

[4] Öhman, S. (1966), Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America 39,* 151-168.

[5] Vilain, A., Abry, C., Badin, P. (1998), A propos des degrés de liberté dans la coarticulation d'un locuteur français filmé 50 fois par secondes, *XXIIè Journées d'Etudes sur la Parole*, Martigny, Suisse, juin 1998

[6] Vilain A., Abry C., Badin P. (1998), Coarticulation and degrees of freedom in the elaboration of a new articulatory plant: Gentiane, *Vth International Congress on˝ Spoken Language Processing'98,* Sydney, Australia, December 1998.