

COPRODUCTION STRATEGIES IN FRENCH VCVS: CONFRONTING ÖHMAN'S MODEL WITH ADULT AND DEVELOPMENTAL ARTICULATORY DATA

Vilain, A., Abry, C., Badin, P.

Institut de la Communication Parlée(ICP) - CNRS UPRESA 5009

Université Stendhal, BP 25, Grenoble cedex 9, France

e-mail: avilain@icp.inpg.fr

ABSTRACT

Articulatory models give a valuable insight to the actions of the degrees of freedom of the vocal tract, thus allowing to test efficiently the classical theories of coarticulation. In this perspective, we use an anthropomorphic articulatory model of a French speaker to investigate compensation strategies in the production of VCVs. The results finally backup Öhman's dual-channel model of coarticulation, and, at a higher level, give further evidence for a dual-channel motor control of speech.

We then question the applicability of this model to the issue of speech ontogeny. Recent developmental studies lead us to propose that the unfolding of speech acquisition goes from (i) the mastering of the carrier articulator, i.e. the mandible, that appears under the form of canonical babbling, to (ii) the independence of the other articulators which allows the control of *local* constrictions, therefore filling the previous proto-syllabic frames with segmental content, then (iii) the learning of *global* control of the vocal tract, maybe guided by auditory input from the environment, and the tuning of coarticulation strategies to adult norms, during the first years of life.

1 INTRODUCTION

As compared to direct observation of raw cineradiographic data, articulatory models represent a more efficient way to disentangle coarticulatory processes in speech, to make out the constraints and variability that apply in the production of each phonetic type, hence to question the classical models of coarticulation.

The models used in this study have been designed in such a way that their command parameters represent the degrees of freedom of the vocal tract of the speaker. This modelling method allows, conversely, to study the actions of the degrees of freedom, through the model, and, at a higher level, to infer motor commands from the observed behaviours.

Our analysis of the coarticulatory behaviour of a French speaker, with the help of an articulatory model, focuses here on the paradoxical example of an [aba] sequence, that seems to present lingual compensatory strategies contradictory with previous results.

We then question the ability of speech production models to describe the acquisition of speech, from the first syllable-like utterances of canonical babbling to the complex global control of coarticulated CV syllables.. Recent developmental data are used to look for a valid scenario for the setting-up of coarticulation in speech ontogeny.

2 MATERIAL AND METHOD

2.1 A statistical articulatory model built from cineradiographic data

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 disclosed with an articulatorily-driven Principal Component statistical Analysis and used as command parameters of the model.

These nine parameters, namely Jaw Height (JH), Tongue Body (TB), Tongue Dorsum (TD), Tongue Tip (TT), Lip Height (LH), Lip Protrusion (LP), Lip Vertical Elevation (LV), Larynx Height (LH), and Tongue Advancement (TA), can be assumed to represent fairly well the degrees of freedom of the articulators of the vocal tract.

The present study is based on one of these two models.

2.2 Extracting the individual actions of the degrees of freedom

We elaborated a method to analyse the coarticulation patterns implemented for the production of the VCV sequences in our corpus (it is described extensively in [Vil99]). The model allowed us to have a view of global vocal tract contours, but also an insight to the individual actions of each degree of freedom of the vocal tract, not only those of the jaw and lips, which are quite visible, but also those of the tongue, namely Tongue Body, Dorsum and Tip. That is we decompose the sagittal function into the different activated articulatory parameters and follow the evolving recruitment of these parameters from the vowel into the consonant, for example.

The bottom figures hereafter exemplify this way of extracting the actions of the first 4 articulatory parameters: The horizontal axis gives the lines of the midsagittal grid used to measure and model the vocal tract, from the glottis (line 1) to the apex (line 27). The vertical axis gives the sagittal distance. The upper solid line is the external contour of the vocal tract, from the pharyngeal wall, to the palate and the teeth. The lower solid line is the neutral position, to which has been added the variation due to the actions of the last five command parameters of the model: actually only two of them do have an effect on the tongue, namely Tongue Advance (horizontal displacement of the tip of the tongue), and Larynx Height, the other 3 parameters concerning the Lips (Lip Height, Lip Protrusion, Lip Vertical Elevation). Our analysis neglects the observation of these parameters, because their effects are very small as compared to those of the first four ones.

The "circled" line represents the sagittal position of the vocal tract, once the action of the JawHeight parameter has been added. Then the action of the Tongue Body parameter is added, which yields the position of the "plus" line; then the Tongue Dorsum gives the "x-marked" line. The addition of the Tongue Tip action brings the final vocal tract configuration, the "star" line, all parameter effects having been computed.

This successive addition of the effects of the parameters shows *a fortiori* when these effects counteract (e.g.: when the action of JH should have lifted up the tongue, but TB brings it back down), therefore highlights the compensation strategies that are implemented in coarticulated speech.

3 COARTICULATION MODELS QUESTIONED BY A PARADOXICAL SAMPLE

3.1 Focus on a puzzling sequence

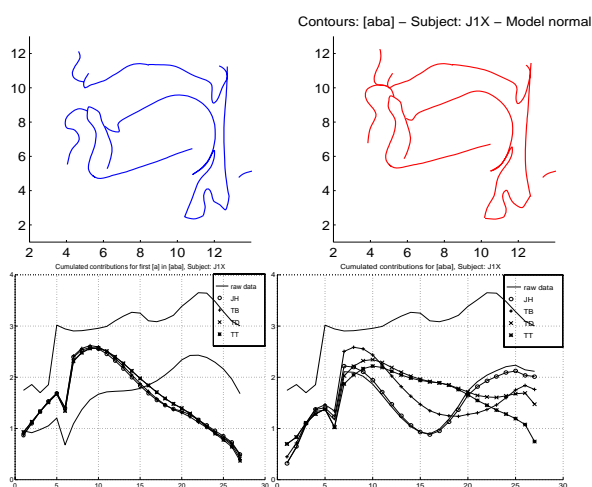


Figure 1: First [a] and [b] in an [aba] sequence produced by speaker J1X: sagittal contours (top figures), and decomposition of the sagittal function into the first 4 command parameters, JH (o), TB (+), TD (x), TT (*) (bottom figures).

Our analysis focused on the example of the [aba] showed in Figure 1. The sagittal contours in the top figures show that the open configuration of the [a] vowel is still present during the performing of the labial occlusion. What is this position due to? If we now observe the bottom figures we can see how the vowel and consonant are completed. The first [a] is realised with the sole action of the mandible, since the effect of the JawHeight parameter symbolised with the circled line, can be seen explaining almost all of the variation from the neutral position. The next [a], that comes after the consonant, has exactly the same articulatory characteristics as the first one. Now the completion of the consonant is made with quite a different strategy: JawHeight is now recruited for another job, i.e. helping the lower lip reach the occlusion. And the noticeable fact is that it is now the tongue command parameters that are acting to keep the vocal tract anterior opening, hence the back constriction.

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.

To illustrate the extent of this phenomenon, we have simulated in Figure 2 what would happen if, as we first thought it would, the consonantal gesture recruiting the jaw and lips only, was simply superimposed on the previous vowel configuration. The simulated contour is in solid line, and the original is dashed.

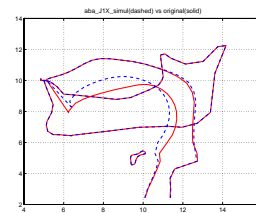


Figure 2: Comparison between the original [b] configuration (solid) and a simulated configuration without the compensating activity of the tongue (dashed)

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.

It is to be noted that the [ava] sequence in our corpus shows exactly the same pattern.

3.2 Questions

This unfolding of events raises questions: (i) Is this lingual activity due to a rapid change in motor commands?; (ii) Does this lingual activity pertain to the vowel, i.e. does this counteract 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? If so, the vowel would be controlled through two different articulatory strategies, between each of which we would have to switch, from vowel to consonant realisations ...an abrupt change which seems unlikely; (iii) Or does it pertain to the consonant in which case these different strategies represent the successive activation of the vowel and consonant control?

3.3 A phoneme-to-phoneme process?

One possibility is that the observed pattern is due to a context free, phoneme-to-phoneme activation process, such as Joos's overlapping innervation wave theory proposes [Joo48]. But the admittance of such a process implies that the lingual gesture should be considered as a constraint pertaining to the labial consonant, in the same way as the jaw raising and the labial upward movement. This solution goes against the idea of labials being unspecified for lingual gestures, and cannot be explained by any articulatory or acoustic need.

3.4 Öhman's model at the motor control level

In the original theory of coarticulation of Öhman [Öhm67], the dual-channel model is explained by the fact that tongue activity in speech is controlled by three sets of muscles with separate neural representation: one for vowels, one for apical consonants, and one for dorsals. The articulatory commands are therefore transmitted independently of each other, which

allows the diphthongal vowel-to-vowel gesture to be carried out while the consonantal gesture is superimposed on it.

What can be proposed here is that the configuration for the [a] vowel can be completed with the only action of the jaw, but that actually the lingual muscles are also recruited during the jaw movement. The activation may not result in an effective displacement, but it still exists. And it is the recruitment of the jaw for the consonant production that lets appear here the tongue muscle activation. The control for the [a] vowel, that was supposed to consist in a jaw opening alone, is here revealed as more complex, since it involves the tongue as well.

This solution is in keeping with the Öhman model of coarticulation at the level of motor commands, in the sense that the vowel gesture is actually constant at the control level, and the fact that the consonant is superimposed on this permanent command lets the activation emerge under different forms.

What has been added therefore is that the motor commands corresponding to a phonetic form may not be effectively operating for one realisation of this form, yet they are actually activated at a high level, and may be revealed by a perturbation, here the consonant gesture.

How do we develop such a complex control over coarticulatory patterns? Are we endowed with this capacity from birth or is it to be acquired in the first years of life? Some studies help us make out the unfolding of this acquisition process.

4 WHAT DEVELOPMENTAL PATH FOR COARTICULATION ?

4.1 Emergence of local constriction control

Munhall & Jones [Mun98] have recorded the activities of the lower and upper lips during the production of [bababa] sequences, by an 8-month old baby and an adult (Figure 3). The baby pattern shows the oscillation of the lower lip carried by the mandible and lower lip from closed to open configuration, but no active movement of the upper lip. "The only upper lip movement occurs in phase with the lower lip motion and is presumably caused by the lower lip forces pushing the upper lip upward after contact."

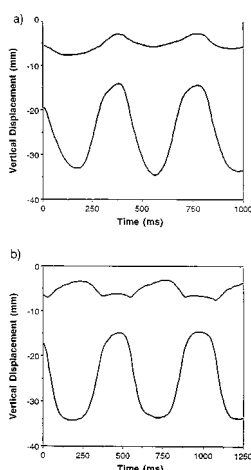


Figure 3: Vertical displacement of upper lip (top trace in each figure) and lower lip (bottom trace) as a function of time during repetitive production of [bababa] by (a) an 8-month-old and (b) an adult.

In contrast, the adult pattern shows a similar oscillation of the lower lip together with the jaw, but also an active displacement of the upper lip, that moves downward to meet the lower lip, and then resists against the pushing.

These two patterns exemplify the acquisition of the control of the lips for speech. This local control, that helps completing precise constrictions is obviously not yet present in the 8-month old baby, i.e. by the time of canonical babbling, and becomes available later in the development.

What then of the global control of the vocal tract that is necessary for the realisation of coarticulated vowels and consonants ?

4.2 Global control strategies come next.

Sussman & al [Sus99] have studied the development of CV coarticulation through the acoustic products, that is with the help of locus equations. Their case study follows a female baby from age 7 months to age 40 months, i.e. from babbling to meaningful speech, and analyses its utterances to obtain F2 at onsets and at vocalic centres. The locus equations are used as an index of the evolving degree of coarticulation of labial, alveolar, and velar consonants across different vowel contexts from baby to adult patterns, taken as representative of the progressive acquisition of segmental independence and the mastering of coarticulation.

The articulatory paths towards adult-like norms appear to be different across consonant types. The labials, for which the baby has to learn to "integrate the vowel tongue position with and independently of the lip closure", show slow slopes at first, that rise towards adult target about 14 months, probably indicating that the child is beginning to produce vocalic tongue placements independent from the consonant.

"Coarticulatory effects" in alveolars start with a high slope, which decreases rapidly in the first months, and then remain under adult-scores until the third year. It can be hypothesised that the child is at first unable to produce independent tongue movements with the two parts of the tongue needed for alveolars and vowels. Then she gains more independent motor control over the different degrees of freedom of his tongue, probably guided by the auditory input she receives from her environment. And then, the very low slopes during the next months seem to indicate a behaviour of hyper-resistance to coarticulation, which is a classical phenomenon in the acquisition of new skills. The newly acquired capacity for independent control tends to be over-exploited, before being finely tuned according to the actual needs of communication.

Velars show the most stable slope values relative to the adult norms: after large fluctuations in the very first months, they keep a high level of coarticulation, even higher than the adult score.

As a matter of fact, it seems easier to produce CVs where both consonant and vowel share the same articulator (velar+vowel), therefore they overlap completely, than to "integrate (labial+vowel), or differentiate (alveolar+vowel) articulatory components in the achievement of CV utterances that meet adult norms."

What appears here, is that when the baby possesses the ability to produce differentiated consonant constrictions, she still has to learn global vocal tract control —a la Öhman, that is from the larynx to the lips—, to adapt her production strategies to the adult coarticulation norms, norms that are specific to her mother tongue, a skill that takes several years to be acquired.

5 CONCLUSION : A LOCAL-TO-GLOBAL SETTING UP OF SPEECH

We have, in a previous paper ([Vil99bis]), tested Peter MacNeilage's notion of babbling consisting initially in rhythmic Frames, entirely attributable to oscillations of the mandible, the baby not being able yet to control its other articulators for speech (for MacNeilage's theory, see [Mac99]). These Frames are then filled with segmental "Content", as the baby gains more independence of its articulators.

Our simulation showed that a mandibular oscillation with articulatory models will yield the same products as those found in babbling, either labial or alveolar constrictions depending on the speaker's morphology, which supported the "Frames, then Content" theory.

Munhall's recordings of the passive lip movement in the 8-month-old brings yet another evidence for this behaviour: "This pattern is consistent with [MacNeilage's] proposal that initial babbling primarily involves mandibular motion.". No control over lip constriction for speech has yet emerged.

Therefore the development of speech coarticulation seems to proceed as follows:

- (i) the first control is that of the mandible, that is the establishment of proto-syllabic Frames, the mastering of the rhythm. Neurophysiological data give evidence that this precociously emerging control originates in the Medial Premotor System, with the Supplementary Motor Area as the speech pacer and initiator (see [Abr00]);
- (ii) as independence of articulators is gained, segmental Content emerges, i.e. the baby begins to control constrictions, e.g.: the control of the upper lip, obviously not yet mastered at 8 months, as seen before, will be available a few months later, to allow the control of more precise local constrictions (a study is currently being carried out in our lab to determine the onset of this mastering, by our colleague S. Brosda);
- (iii) Sussman's study shows that the global control of the vocal tract needed to perform adult-like coarticulation patterns, as the one described in 3., is only coming later in the development.

So speech develops from Frames, to Content, then to coarticulated Content.

6 BIBLIOGRAPHY

- [Abr98] Abry, C., Boë L.-J., Laboissière R., Schwartz J.-L. (1998), A new puzzle for the evolution of speech ?, Peer Commentary to [Mac98], , *Behavioural and Brain Sciences*, 21, 4, pp.512-513.
- [Abr00] Abry, C., Stefanuto, M., Vilain, A., Laboissiere, R. (2000), "What can the utterance "tan, tan" of Broca's patient Leborgne tell us about the hypothesis of an emergent "babble-syllable" downloaded by SMA?", in *Phonology, from phonetics to cognition* , J. Durand & B. Laks Eds, Oxford University Press, to be published.
- [Bro99] Brosda, S. (1999), "De la variation dans le babillage canonique: l'apprentissage sensori-moteur", DEA de Sciences Cognitives, ICP, Grenoble.
- [Joo48] Joos, M. (1948), Acoustic Phonetics, *Journal of the Linguistic Society of America*, vol.24, n°2, suppl.

[Mac98] MacNeilage, P. (1998), The frame/content theory of evolution of speech production, *Behavioural and Brain Sciences*, 21, 4, pp. 499-521.

[Mun98] Munhall, K.G., Jones, J.A. (1998), Articulatory evidence for syllabic structure, *Behavioural and Brain Sciences*, 21, 4, pp.524-525.

[Öhm67] Öhman, S.E.G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, 41, 310-320.

[Sus99] Sussman, H., Duder, C., Dalston, E., Cacciatore, A. (1999), A acoustic analysis of the development of CV coarticulation: a case study, *Journal of Speech, Language, and Hearing Research*, 42, 1080-1096.

[Vil99] Vilain A., Abry C., Badin P. (1999), Motor equivalence evidenced by articulatory modelling, *Eurospeech'99*, Budapest, Hungary.

[Vil99bis] Vilain A., Abry, C., Badin, P., Brosda, S. (1999), From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling, *ICPhS'99*, San Francisco, USA.