MODELING LIP CONSTRICTION ANTICIPATORY BEHAVIOUR FOR ROUNDING IN FRENCH WITH THE MEM (*MOVEMENT EXPANSION MODEL*)

C. Abry and T.M. Lallouache
ICP URA CNRS n° 368 INPG/ENSERG Université Stendhal
BP 25 F-38 040 Grenoble Cedex 9

ABSTRACT

A new model called MEM, *Movement Expansion Model*, is proposed as an alternative to current anticipatory models. Initially developed to deal with one of the correlate of vocal-tract lenghtening (upper lip protrusion), this model is presently extended to the other main component of rounding, the modulation of between-lips area, which time course has never been integrated in the frame of anticipatory models, in spite of its crucial role in acoustics.

1. INTRODUCTION: Protrusion MEM

We are currently developing a *Movement Expansion Model* (MEM [1]), as an alternative to other models available in the field of speech anticipatory behaviour: the so called *look-ahead* [LA], *time-locked* [TL] (now *frame* or *coproduction*) and *hybrid* ("LA+TL") models [2].

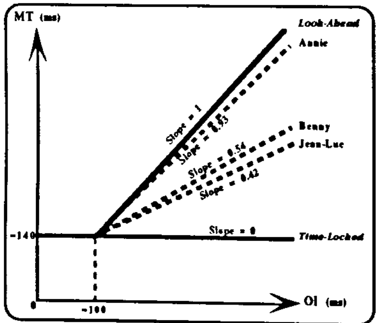

Fig.2– Movement expansion model (MEM): dotted speaker specific regression lines and LA & TL predictions (from [1], see text).

In this model – which classically dealt specifically with upper lip protrusion – in French [i(CCCCC)y] transitions [1], movement time MT was shown to be dependent on the effective duration of the string of consonants (obstruence interval *OI*) produced in the transition from vowel to vowel. We showed that this movement expansion was linearly related to the duration of *OI*, the slope of this relation being speaker-specific. MEM specified for each subject a basic duration for protrusion *MT*, typically 140 ms for [iy] and [iCy] (Fig. 1), as well as an expansion function, starting from about 100 ms *OI* for [iCy]. Speaker-specific parameterization is clearly not in favour of a generalization of either the LA nor TL models or current modified versions. For short, in our [i] to [y] transitions, the anticipation of the protrusion movement is not determined by the end of the unrounded vowel [i], like in LA: in Fig. 1, only one subject, Annie, displays such a behaviour, with a slope near 1. Neither is anticipation determined in a fixed way in relation to the acoustic onset of the rounded vowel [y], like TL: no subject had a relatively constant *MT*, i.e. zero slope was not observed for our other two speakers, Jean-Luc and Benny, whose coefficients are about 0.5.

2. NEED FOR EXTENSION: Constriction MEM

The purpose of the present study was to extend this result to the description of the time course of between-lips area. This area parameter is known to be the most responsible for acoustic changes [3]. It has, to our knowlegde, never been integrated into anticipatory models and other students in the field have called for it [2]. Since we have the image processing system to measure accurately this parameter [4], it was planned for us since the beginning of our work that we would have to use it within an anticipatory model. We were only refrained to do so by the inspection of these lip-area temporal functions. These are rather "bumpy" due to the action of the jaw recruited to produce coronal consonants, like [s,t,l]: the elevation of this carrier articulator diminishes the area, without any active movement of the lips.

In addition, in [1] we could use only 3 out of the 4 subjects we recorded initially, since one of them (in spite of being French !) displayed quite no upper lip protrusion, except a few 10th of millimeter, compared to a range of about 8 mm for the others.
We will show that when we use main events to describe such area profiles, it is possible to predict the time course of the constriction of the vocal-tract output with the same MEM model we used for protrusion.

3. METHOD

Processed face signals (27,000 frames) were the same as in [1], with transitions ranging from [iy] to [ikstsky]. Labelling of audio and video signals was also the same for *OI* interval and kinematic events.
The procedure used to detect events on temporal functions of between-lips area for [i] to [y] transitions was specially designed to maximally avoid small consonantal perturbations (hence ambiguities). Since the obtained curves reflect pretty accurate measurements, it was not chosen to smoothe systematically such perturbations and the weight used for cubic splines to get a continuous function was high enough. So kinematic events used for protrusion and obtained from derivatives were discarded, being too sensitive. We finally characterized these movement profiles with 5 events. First, we considered that when a 10% value of Area Amplitude (10% [Max.Area-Min.Area]), was reached (*10%Area.On*[set]), held or diminished, then increased (*10%Area.Off*[set]), we could safely determine a "Hold" (*H*) phase where acoustic efficiency of constriction was ascertained enough. It follows that we detect *90%Area.On*, reflecting the onset of the constriction movement towards [y], and of course *Max.Area* and *Min.Area*. The "Time Falling" (*Tf*) phase begins with *90%Area.On*, and ends with *10%Area.On*. We will finally, among other phase combinations, use *Tf+H* as a global phase to get the best overall prediction of movement expansion.

4. RESULTS

4.1. Constriction phases and *OI*

Taking advantage of the procedure we used for the study of upper lip protrusion, we chose not to begin by the examination of articulatory events referenced to the acoustic domain, but we searched for correlations between the two flows, i.e. articulatory phases with OI, without a common reference event (in order to avoid part-whole correlation artefact [5]).
Correlation coefficients ($r = 0.32$ at $p = 0.01$) were calculated for all phases without [iy] and with [iy] in order to test intercept values, since the test of the MEM (contrary to other procedures used currently by other students [2]), allows to evaluate, from all samples where *OI* is different from zero, the prediction of the basic simple transition gesture duration.
Figs. 2a-d show the piecewise fitting for each speaker in *Tf+H* and *OI*. Notice that there is no real temporal continuum between one-consonant sequence and the others. Other prosodic factors should certainly be manipulated (for example rate) to cover the whole range of variation of this obstruence interval (*OI*), variously filled, depending on the habits of each speaker. r values corresponding to calculations without [iy] are all very high (from 0.87 to 0.99), contrasting sharply, like for upper lip anticipation, with quantitative data published for English [2]. As in the case of upper lip, intercepts given by these linear regressions cannot predict accurately enough the mean duration of the simple gesture (Jean-Luc: 129 *vs.* 158 ms; Annie: 90 *vs.* 161 ms; Benny: 85 *vs.* 148 ms; Christophe: 67 *vs.* 107 ms) and so the piecewise fitting reveals generally more appropriate.
If we consider now the slopes, it is also clear that only one speaker (Annie, the same as for upper lip protrusion) approaches the LA model (with 0.93), the three others behaving in rather close individual range (between 0.69 and 0.79), higher than for upper lip slopes (Fig. 1), but still not in the orthodoxy of LA (not to speak of TL).
If we want to give a schemata of these results, the only main difference with upper lip protrusion behaviour (Fig. 1), stays simply in the fact that the newly processed speaker (Christophe) has a rather small 100 ms duration for his basic contriction gesture [iy], compared with the 150-160 ms durations for the three others. But this is not a problem for our

model since the MEM specifies for each speaker his basic gesture values, then calculates every expansion knowing his expansion function, that can be fairly obtained with some test sample, manipulating *OI* from one consonant (about 100 ms for all speakers) to three or more (the maximum *OI* value depending on each speaker's rate habits: under 300 ms or up to about 400 ms).

4.1. Constriction phasing in *OI*

It is time now to set these results in relation to the acoustics, chosing a common reference event, to test if the procedure we used for upper lip protrusion is viable for lip constriction. To make short we will give only one example, Jean-Luc, knowing that the expansion functions we gave on Fig. 2 offer the possibility to calculate the fitness of the data of each speaker [1].

In Fig. 3a we represented, for this speaker, only the upper lip protrusion kinematic event *PO* (for *Protrusion Onset*), with the offset of [i] (*VVT[i]*) as the reference event (lower horizontal line at 0%; onset of [y], *VVO[y]* is the horizontal line at 100%). *%(PO-VVT[i])/OI* provides thus a *relative timing* measurement, say *phasing*.

In Fig. 3b, we did the same for area changes, using *10%Area.On* as a landmark comparable to *PO*, with the same reference event *VVT[i]*.

How does anticipation of these two events behave ? For movement onset (*PO*), it is clear that data point dispersion adopts a hyperbolic function (for this speaker as for others high regression coefficients were obtained with this fitting, from 0.82 to 0.93 [1]). The onset of protrusion can occur relatively well *into* the vowel [i] for small *OI* values (one consonant); and clearly *outside* of it (for *OI* values above 300 ms, corresponding here mainly to five consonants). We observe the same trend, with relatively less amplitude, for the constriction beginning event *10%Area.On*.

So we can say that the MEM holds for the two main components of rounding, protrusion and constriction.

5. DISCUSSION

Our Movement Expansion Model succeeds in accounting for the behaviour of the four French speakers under examination. MEM specifies for each a basic duration for the protrusion and constriction components of rounding, as well as an expansion function with a speaker-specific parameterization.

The fact that expansion coefficients vary between speakers may be reminiscent of a more abstract view of variation in language, i.e. the so-called "principles and parameters" approach in Chomsky's Universal Grammar. But in our concrete measurements this means simply that subjects follow globally and coherently the same expansion "law", with subject-specific parameters.

So to speak: vocalic gestures expand when they have temporal room enough between each other, regularly and at each speaker's own rate, without any "obligatory principle" urging them to fill between-vowel interval. This is a fairly different conception from both the look-ahead and time-locked ones.

Further work is in progress to test the MEM with the two other main components of vowel gestures: high-low and front-back dimensions.

Acknowledgements: This work was done in the frame of Esprit Basic Research project n°6975 *Speech Maps*.

REFERENCES

[1] Abry C. & Lallouache M.T. (in press). Pour un modèle d'anticipation dépendant du locuteur. Données sur l'arrondissement en français, *Bulletin de la Communication Parlée*.

[2] Perkell J.S. & Matthies M.L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/ : Within- and cross-subject variability, *Journal of the Acoustical Society of America, 91*, 2911-2925.

[3] Badin P., Motoki K., Miki N., Ritterhaus D. & Lallouache T.M. (1994). Some geometric and acoustic properties of the lip horn, *Journal of the Acoustical Society of Japan (E) 15, 4*, 243-253.

[4] Lallouache M.T. (1991). *Un poste «Visage-Parole» couleur. Acquisition et traitement automatique des contours des lèvres*, Thèse de l'INP, Grenoble.

[5] Benoît C. (1986). Note on the use of correlations in speech timing, *Journal of the Acoustical Society of America, 80*, 1846-1849.

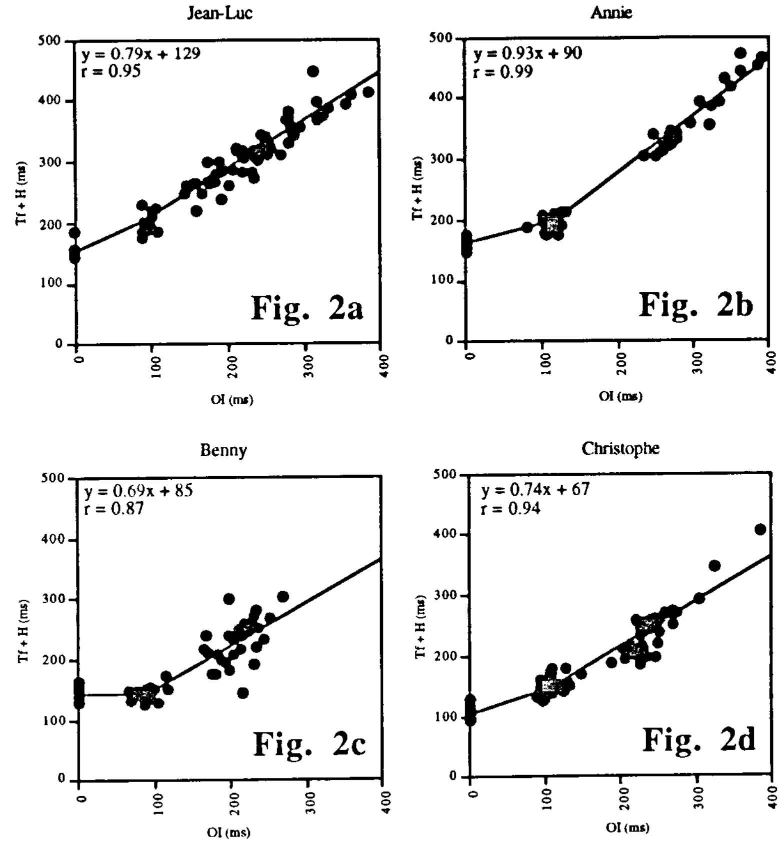

Fig. 2a Fig. 2b Fig. 2c Fig. 2d

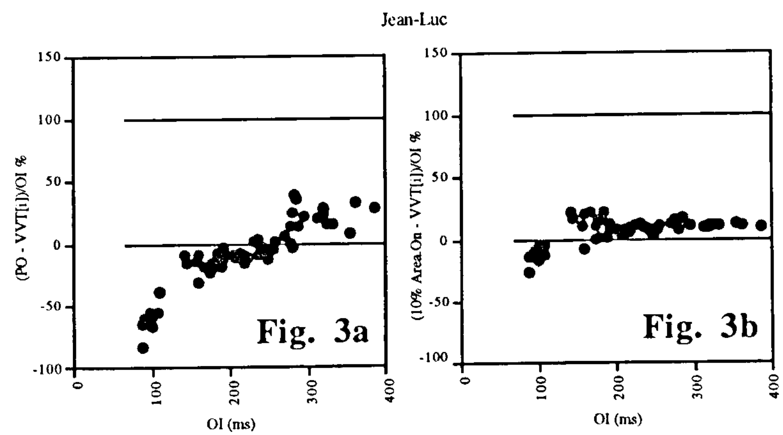

Fig. 3a Fig. 3b