The development of anticipatory labial coarticulation in French: A pionneering study

Aude Noiray [a], Lucie Ménard[b], Marie-Agnès Cathiard [a], Christian Abry [a] & Christophe Savariaux [a]

[a] Institut de la Communication Parlée, Université Stendhal-INPG, UMR 5009, 38040 Grenoble cedex 9, France
[b]Département de linguistique et de didactique des langues, Université du Québec à Montréal, Case postale 8888, succursale Centre-ville, Montréal (Québec) H3C 3P8, Canada
noiray@icp.inpg.fr menard.lucie@uqam.ca

Abstract

This article reports an experimental study initiated on labial anticipatory coarticulation in the framework of a description of motor control development. Four French children between 4 and 8 years old have been audio-visually recorded, uttering [iC$_n$y] puppets names, in which C$_n$ corresponds to a varying number of intervocalic consonants. A kinetic lip area function has been obtained via the ICP tracking system (Lallouache [1]) in order to describe the anticipatory movements of vocalic targets. As several concurrent models have been evaluated to account for anticipation in adults but very few for children, and since the more robust has proved for French to be the *Movement Expansion Model* (MEM; Abry & Lallouache [2], [3], [4], Abry *et al.* [5]) we adopted this framework for testing the anticipatory motor behaviour in children.

1. Introduction

The dynamic control of supralaryngeal articulators recruited for coarticulation as well as anticipation planification is not adult-like until mid-childhood (Kent [6], [7]). Thus, it is a crucial point to study how these coarticulatory strategies develop through time: is there a critical age for the acquisition of labial anticipatory mechanism, or does it emerge progressively?

The literature often refers to three protrusion anticipation models for adult speakers, proposing a different interpretation of the temporal span for the labial rounding movement. On the one hand, Henke ([8] for English), Benguerel & Cowan ([9], for French) and Lubker ([10], for Swedish) are proponent of a rounding movement expansion proportional to the consonantal interval, starting at the acoustic offset of the unrounded vowel [i] (*Look Ahead Model*: LA). On the other hand, Bell-Berti and Harris [11] are in favour of a time locked anticipation movement whatever the number of intervocalic consonants *(Time Locked Model*: TL). According to the authors, the articulatory movement onset remains independent from the length of the previous phone string and always starts at a fixed time before the acoustic onset of the rounded vowel. A third model has been proposed in response to data in contradiction with the two models described above (Perkell & Chiang [12]). According to the *Hybrid Model* the protrusion movement is divided into two steps: an initial slow phase starting like in the LA model at the offset of the unrounded vowel, followed by a more rapid phase, with a time-locked acceleration peak. Therefore, the more the intervocalic string increases in duration, the more the initial slow phase expands.

During the two past decades, experimental studies have been run, standing up for either of the three models. Yet, one of the major study carried out by Perkell ([13], Perkell & Matthies [14]) on English adult speakers highlighted a great variability in results for protrusion anticipation. The results obtained for the four speakers investigated displayed for protrusion movement scattered data in between TL (zero slope) and LA (slope = 1) for each speaker. Hence Perkell ([13], p.280) concluded that these "data […] allow us to reject strong versions of all three models".

An alternative model we called MEM, *Movement Expansion Model* has been elaborated for French by Abry and Lallouache ([2], [3]). An experimental study on four French adult speakers suggested that for protrusion movement time (MT): (1) there is an incompressible duration (*execution constant*) of about 140 ms for the sequence [iy], which is not systematically different from [iCy] ; (2) from [iCy] to [iCCCCy] there is a linear expansion of MT; (3) the correlation coefficient are all clearly significant: hence the scattering in between TL and LA is not the case; (4) the slopes of the regression lines are *speaker specific*; (5) since slopes range from 0.42 to 0.93, LA is not general but just one possible extreme speaker-specific behaviour. We can notice that the MEM is not limited to lip protrusion parameters, i.e. Protrusion MEM: Constriction MEM captures also the kinetic lip area function. Hence it can account even for a speaker having too little a protrusion amplitude to be processed (Abry & Lallouache [4] [5]; for a reception of this model, see Farnetani & Recasens [15]; and recently Byrd and Saltzman [16]).

As concerns children, these models have not been comparatively tested. Abelin *et al.* [17] conducted an experimental study on anticipatory lip rounding behaviour of six Swedish children, from 7 to 10 years old, with one adult as a reference (the one investigated by Lubker [10]). EMG activity was recorded and processed for detecting rounding onset. Only one child displayed the typical pattern corresponding to the LA behaviour evidenced by Lubker for the reference adult speaker (with a slope around

1). The five remaining children showed rather scattered results, which contrary to the author's claim cannot be related confidently to either LA or TL models.

2. Method

2.1. Subjects and corpus

Four French children (three boys KV:4,5; DV:8 and EC:4 years old and a girl AC:4), without schooling problems, and two female adults (CL and MA) have been recorded. None of them had hearing or articulatory disabilities; no perceptible French dialectal accents were noticed. The task consisted in repetitions of a series of puppet names prompted by the experimenter, typically [i(C_n)y] (in which C_n corresponded to a number varying from 0 to 3 intervocalic consonants). These names ([iy], [isy], [iky], [iksy], [ikry], [itkry], [iskry], [iksty]) were integrated into sentences like « le toutou Iku est rouge » (the doggie Iku is red).

Because of the age of our young speakers, a real effort has been made for the preparation of these recording sessions so that the children felt confident and interested in the experiment. They were told it was a work on characters voices in comics. They were first trained to play with the eight puppets corresponding to the eight names. This preliminary step was used to familiarize children with experimenters, and to make them easily memorize the names of the puppets. To render the sound booth more a friendly place for children, it was decorated with sheen material, fairy lights, luminous stars and puppets. At the end of the recording, a snack was organized and each children was given a present.

2.2. Audio-visual recording

Each speaker has been audio-visually recorded in the anechoic chamber with the ICP tracking system [1]. The images were recorded on a Betacam tape via two cameras (face and profile). The images were digitized at a 25Hz rate. To minimize face or body movements, a baby seat and a piece of foam rubber slightly dug for the children head was used. Since two 5 500 watts fluorescent neon lights had to be used, children wore black glasses for protection. Two calibration blue discs were stuck on the glasses and a small ruler painted in blue was fixed hanging on one side-piece so that profile references could be obtained. Lips were made up with the same waterproof saturated blue (lipstick). This blue was chosen in order to facilitate the acquisition of labial contours with an automatic detection system developed at ICP via numerical Chromakey processing. Since one video image corresponds to two interleaved frames, it was possible to obtain a 50 Hz rate through line interpolation, synchronised with the digitized acoustic signal (44.1 kHz). Lateral (protrusion) and front lip shape and position parameters were automatically detected for each sequence with Tacle software (Audouy [18]).

2.3. Articulatory analysis

In order to study rounding anticipation, we will focus here on the major moving shape, i.e. between lip area (constriction), which is relevant both visually and acoustically. Other events on velocity and acceleration derivatives have been studied in a preliminary report on 2 of the children and the 2 adults here investigated (Ménard et al. [19]). In accordance with the analysis led by Abry & Lallouache [4], they do not change the conclusions in favour of the MEM model. Note that the area parameter remains tricky to detect since it is a shape processing and not a flesh-point tracking. In addition to classical kinetic events, robust landmarks can be detected on this constriction temporal function using the classical 10% and 90% levels describing *Time Falling* and *Time Rising* on hysteresis curves. Therefore, according to previous studies and in the framework of the MEM, several acoustic and articulatory events have been detected (Abry *et al.*, [5]). First, the obstruence interval (IO) was determined on the spectrogram, *i.e.* the duration between the acoustic offset of the vowel [i] (characterized by the disappearance of the upper formants structure) and onset of [y] (by formants appearing again). On the time course of lip area, the maximal area for [i] (event 1 on figure 1) and minimal area for [y] (event 4) were measured. Three other events were identified: time corresponding to a 10% decrease of area amplitude (event 2), time corresponding to 90% of this range (event 3), and finally time corresponding to a 10% increase of lip area following minimal area of the vowel [y] (event 5). The interval between event 2 and event 3 is *Time Falling* (TF) and the interval between event 3 and 5 corresponds to an *Hold* (H), that is a phase in which acoustic efficiency of constriction is ascertained.

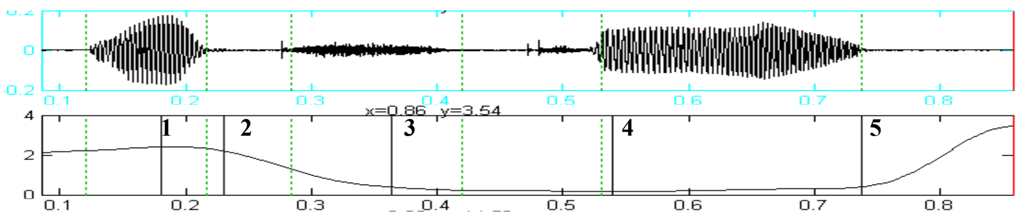

Figure1 : Acoustic signal (above) and lip area time course (below) for a sequence [iksty] uttered by a child. Identification of events : 1= maximal area for vowel [i] ; 2=10% of area difference between [i] and [y] ; 3=90% of area range between [i] and [y] ; 4= minimal area for [y] ; 5=10% of lip area difference between [i] and [y] following minimal area of vowel [y]. Interval 3-2 corresponds to *Time Falling* (TF) and interval 3-5 to the duration of *Hold* phase (H).

3. Results

In order to describe our speakers' anticipatory behaviour in the MEM framework, the total duration of constriction movement (TF+H) was plotted against the obstruence interval (IO) (figure 2). The two adults and two of the children uttered [iy] with a zero IO value, i.e. with no interruption between the two vowels. For these subjects, mean value was calculated, showing that each *execution constant* was quite stable within a range between 350-380 ms. Consequently, the linear regression analysis (correlation coefficient and slope) was performed differently: on all [iCy], [iCCy], [iCCCy] sequences, [iy] excluded, for these speakers; and on all sequences for KV and DV children including [iy] they realized [iʔy] (the horizontal line, drawn from their lowest IO value and intersecting the y-axis, is just here for comparison's sake).
Only one child (KV) had scattered results with a non significant correlation coefficient. The slopes measured for the three other children ranged from 0.63 to 1.22. The two adults MA and CL have respectively 0.60 and 0.81 slopes (table 1). These values indicate that movement expansion is linearly related to the duration of OI, the slope of this relation being speaker-specific (including the "*hyper LA*" 1.22 value). In figure 3, Abry *et al*.'s data [5] obtained for four French subjects on upper lip protrusion and lip constriction (area) are summarized. One of their subjects, "Christophe" had a too small protrusion amplitude to be processed and consequently only his constriction behaviour was available. Therefore, for the Protrusion MEM, the slopes of three subjects ranged from 0.42 to 0.93, whereas the slopes of the four subjects ranged from 0.69 to 0.93 for Constriction MEM.
Taken together these results confirm that the *MEM* can account for labial anticipatory behaviour for 6/6 French adult speakers and 3/4 children. They contrast sharply with Perkell & Matthies' scattered data [14] on protrusion in 4/4 American adults. And also with Abelin *et al*.'s data [17] on EMG labial activity for 5/6 Swedish children.

4. Conclusion

In a language like French (and Swedish, Lubker [10]) for which labial constriction is crucial for the vowel rounding contrast, we found indeed very regular anticipatory behaviour throughout the six adult French speakers we investigated up to now. This means that all speakers display the same significant linear expansion behaviour when the duration of consonants increases in the V-to-V transition, adopting their own idiosyncratic movement expansion rate. In other words, each speaker can adopt a personal slope of movement expansion around which data are fairly grouped, i.e. with a significant correlation coefficient. When comparing this regular behaving adults with children of the same linguistic community, we found that this language specific regular behaviour can be settled very early, as early as the youngest subjects who performed the experiment (about 4 years old). Then it seems that at least for French this type of control can be achieved earlier than Swedish for which Abelin [17] found only one child behaving at 8 like the reference LA adult. However, since one of the youngest French children investigated didn't exhibit this regular behaviour, additional work is needed to focus on this three to five years period.

5. Acknowledgements

We would like to thank Alain Arnal for his technical assistance, the children and their parents for participation. We are also thankful to Virginie Attina & Marion Dohen for their data processing assistance.

6. References

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

[2] Abry, C. & Lallouache, T., "Audibility and Stability of Articulatory Movements: Deciphering two experiments on anticipatory rounding in French", *12th Int. Congress of Phonetic Sciences Proc.*, 1: 220-225, 1991.

[3] Abry, C. & Lallouache, T., "Le *MEM* : un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français", *Bulletin de la Communication Parlée*, Vol.3, 85-99, 1995.

[4] Abry, C., et Lallouache, M. T., "Modeling lip constriction anticipatory behaviour for rounding in French with the MEM (Movement Expansion Model)", *XIIIth Int. Congr. of Phonetic Sciences Proc.*, 4: 152-155, 1995.

[5] Abry, C., Lallouache, M.-T., Cathiard, M.-A., "How can coarticulation models account for speech sensitivity to audio-visual desynchronization?", In D. Stork & M. Hennecke (Eds.), *Speechreading by Humans and Machines,* NATO ASI Series F: Computer and Systems Sciences, 150: 247-255, Springer-Verlag, Berlin Heidelberg New York London Paris Tokyo, 1996.

[6] Kent, R. D., "Anatomical and Neuromuscular Maturation of the Speech Mechanism. Evidence from Acoustic Studies", *J. Speech and Hearing Res*., Vol.19, 421-447, 1976.

[7] Kent, R. D., The Segmental Organization of Speech, In P. F. MacNeilage (Ed.), *The Production of Speech*, New York: Springer-Verlag, 57-86, 1983.

[8] Henke, W. L. "Preliminaries to speech synthesis based on an articulatory model", *IEEE Speech Conf. Proc*., 170-171, 1967.

[9] Benguerel, A.P., Cowan, H.A., Coarticulation of upper lip protrusion in French, *Phonetica*, Vol. 30, 41-55, 1974.

[10] Lubker, J., “Temporal Aspects of Speech Production: Anticipatory Labial Coarticulation”, *Phonetica*, Vol. 38, 51-65, 1981.

[11] Bell-Berti, F. & Harris K. S. Temporal patterns of coarticulation: Lip rounding. *J. Acoust. Soc. Amer.*, Vol 71, 449-459, 1982.

[12] Perkell, J. S. & Chiang, C., Preliminary support for a "hybrid model" of anticipatory coarticulation, *Proc. of the 12th Int. Conf. of Acoustics.*, A3-6, 1986.

[13] Perkell, J.S., Testing theories of speech production: Implications of some detailed analyses of variable articulatory data, In W.J. Hardcastle and A. Marchal (Eds*.), Speech production and speech modeling*, 263-288, Kluwer Academic Publishers, 1990.

[14] Perkell, J.S., L. Mathies. M., “Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability”, *J. Acoust. Soc. Amer.,* Vol. 91(5), 2911-2925, 1992.

[15] Farnetani, E., Recasens, D., Coarticulation models in recent speech production theories, In W.J. Hardscastle and N. Hewlett (Eds.), *Coarticulation : Theory, data and techniques*, Cambridge University Press, 31-68, 1999.

[16] Byrd, D., Saltzman, E., "The elastic phrase: modeling the dynamics of boundary-adjacent lengthening", *Journal of Phonetics*, Vol. 31, 149-180, 2003.

[17] Abelin, A., Landberg I. et Persson. L., “A study of anticipatory labial coarticulation in the speech of children”, *PERILUS*, Vol. 2, 2-18, 1980.

[18] Audouy M., Logiciel de traitement d’images video pour la determination de mouvements des lèvres. Projet de fin d’études, option génie logiciel, ENSIMA, 2000.

[19] Menard, L., Cathiard, M.-A., Savariaux, C., "Le développement de la coarticulation labiale anticipante en français : une étude préliminaire, *Actes des 25èmes Journées d'Etude sur la Parole*, Fèz, Maroc, Avril 2004.

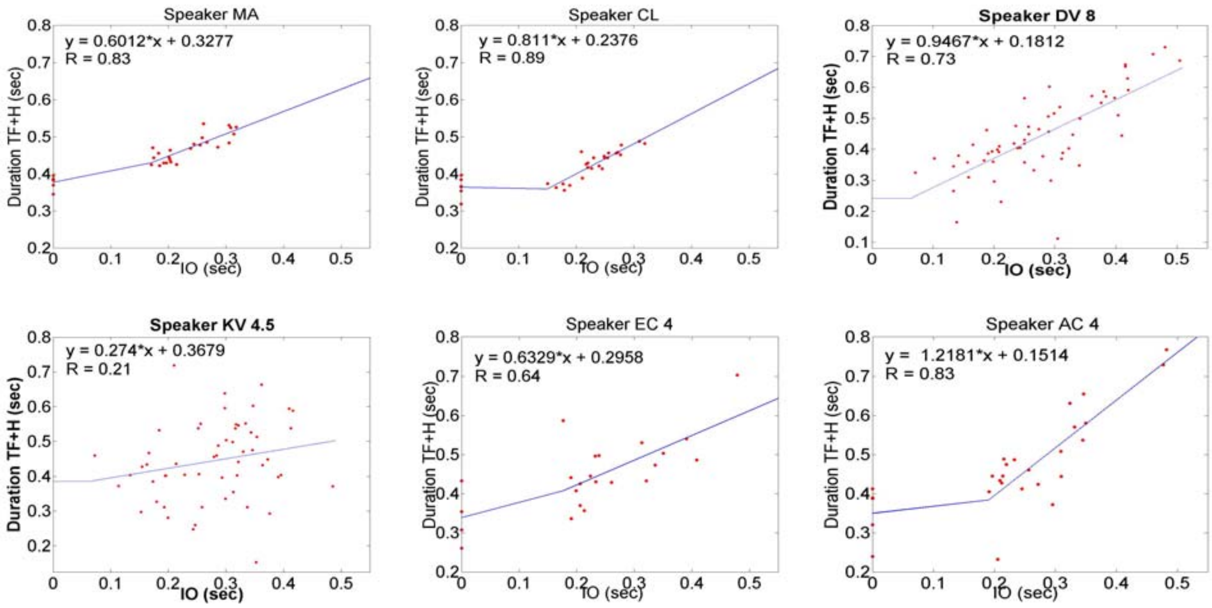

Figure2: Relationships between total duration of lip rounding constriction movement (TF+H) and duration of IO (obstruence interval of intervocalic consonants) in seconds. (For the grouping of the [iC$_n$y] with or without [iy], see text).

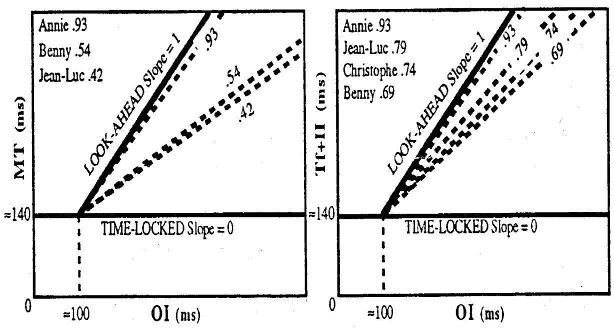

Figure 3: A sketch of the MEM model for protrusion (left) and constriction (right) components (from Abry *et al.*, 1996, fig. 4 & 5).

Speakers	Regression slope	Correlation coeff. (R)
Children		**p<0.01
AC (4)	1.22	0.83 **
EC (4)	0.63	0.64 **
KV (4,5)	0.27	0.21 NS
DV (8)	0.94	0.73**
Adults		
CL	0.81	0.89**
MA	0.60	0.83**

Table 1: Regression slopes and correlation coefficients for the 4 children and the 2 adults of the present study.