Extending the *Movement Expansion Model* (*MEM*) for rounding from French to English

Aude Noiray[1], Lucie Ménard[2], Marie-Agnès Cathiard[1], Christian Abry[1], Jérôme Aubin[2], Christophe Savariaux[1]
[1]Institut de la Communication Parlée, Université Stendhal-INPG, UMR 5009
38040 Grenoble cedex 9, France

[2] Département de linguistique et de didactique des langues,
Université du Québec à Montréal
Case postale 8888, succursale Centre-ville, Montréal (Québec) H3C 3P8, Canada

Aude.Noiray@icp.inpg.fr, menard.lucie@uqam.ca

Abstract. *A test of the current labial anticipatory coarticulation models was performed on 8 adults (4 American English and 4 Canadian French). The subjects were audiovisually recorded, uttering [iC$_n$u] sequences. In order to describe their anticipatory movement patterns, protrusion as well as constriction temporal functions were obtained simultaneously via two motion captures: Optotrak for* flesh-point protrusion, *together with the ICP-Lip-Shape-Tracking-System, which can deliver* area constriction *in addition. A bite-block condition was added in order to control the possible consonantal influence (via the jaw, notably for coronals) on the time course of rounding anticipation. As for French subjects, for whom the* Movement Expansion Model (MEM) *was first elaborated (and retested recently for children), Canadian French display a* lawful MEM *anticipatory behaviour. Surprisingly, this is also the case for American English subjects who have been repeatedly reported not to behave in the line of French or Swedish rounding behaviour. We attribute the failure of the preceding American studies to a too small protrusion magnitude (rare in French but evidenced here for 3 of 4 Americans). The techniques they used could not capture the acoustically relevant lip area parameter, which is now clearly controlled, for French and for American rounding, within a* MEM *anticipatory pattern.*

1. Introduction

In the literature on labial anticipatory coarticulation, three models based on protrusion are classically taken as reference to account for the temporal expansion of the labial rounding movement in adults. The *Look-ahead Model* elaborated by Henke (1966) for English, and tested for French by Benguerel and Cowan (1974), predicts that the rounding gesture would start at the acoustic offset of the unrounded vowel [i], with a movement expansion proportional to the consonantal interval between the two vowels. A diverging interpretation was proposed by Bell-Berti and Harris (1982), who claim that, for a defined speech style, the movement onset is time-locked to the acoustic onset of the rounded vowel whatever the number of intervocalic consonants *(Time-locked Model*). Finally, according to a *Hybrid* model (see Perkell and Matthies, 1992), the rounding gesture would start at the acoustic offset of the unrounded vowel, with an initial slow phase (with a rather *Look-ahead* slope), followed by a more rapid phase associated to a *time-locked* acceleration peak**.**

These three models have been tested in a major study carried out by Perkell and Matthies (1992) on 4 American subjects, for upper lip displacements in [iC_nu] sequences, in which C_n corresponded to a varying number of intervocalic consonants. The variability evidence led Perkell and Matthies to conclude that "*data […] allow us to reject strong versions of all three models*". As an alternative, the *Movement Expansion Model* (*MEM*) has been proposed by Abry and Lallouache (1995). In an experimental study of French adult speakers, they tested the extent of the anticipatory movement on [iC_ny] sequences. The results showed that for protrusion movement time (MT): (1) there was a rather incompressible duration (*execution constant*) of about 140 ms for the sequence [iy], and quite the same for [iCy] ; (2) from [iCy] to [iCCCCCy] there was a linear expansion of MT; (3) the correlation coefficients were all significant; (4) the slopes of the regression lines were *speaker-specific*. Since the slopes of the French speakers ranged from 0.42 to 0.93, a typical *Look-ahead* pattern could not be generalized to all speakers, but only considered as one possible speaker-specific behaviour, that can be described in the framework of *MEM* principles, with a speaker-specific parameterization. We must underscore that, contrary to the variability found by Perkell and Matthies (1992), the variability in our subjects' behaviour was assessed as being *lawful variability*, technically by means of significant correlation coefficients together with regression functions. Since one of the speakers investigated had a too small protrusion to be reliably processed, a *Constriction MEM* was elaborated in order to test the pattern of the rounding constriction of all speakers including the ones having a non detectable protrusion (Abry *et al.*, 1996). More recently, this *Constriction MEM* model succeeded in accounting for rounding anticipatory behaviours of 8 French young children plus two adult controls (Noiray *et al.*, 2004).

The state of the model controversy between American researchers and European ones (since the Swedish have been shown from the beginning to display a non *Time-locked* pattern, Lubker, 1981) was recently formulated by Byrd and Saltzman (2003, note 4, p.157) as a language specific issue. Notice that the discrepancy between French *vs.* American results is fitted in their control model by the use of an *ad hoc* "side constraint" (Rubin *et al.*, 1996). So, is anticipatory labial coarticulation really a language specific control? Or is only one dimension of rounding, namely protrusion, a fragile cue especially for a reputed lack of amplitude in English? Whereas the constriction gesture, in order to achieve acoustically rounding, could be more robust to evidence the anticipatory control pattern?

In the present study, we had the opportunity to design a Quebec French *vs*. American English comparison. In order to describe the subjects' anticipatory movement patterns, protrusion as well as constriction temporal functions were obtained simultaneously via two motion captures: Optotrak for *flesh-point* protrusion, together with the ICP-Lip-Shape-Tracking-System (our so-called "Deep Blue"), which can deliver *area constriction* in addition.

2. Methods

2.1. Subjects and task

Four Canadian French adults (two males: ANT, JER; and two females: ANN, LUC; mean age: 28) and four American English adults (two males: DAN, KAL; and two females: LAU, SAR; mean age: 22) were audiovisually recorded.

The task consisted in the random repetition of a series of [iC_nu] sequences, in order to observe the evolution of their rounding movement time-span, depending on intervocalic consonantal string variation in duration. The rounded vocalic target [u] has been chosen for comparison purposes since it is phonemic in both languages. The resulting sequences ([iu], [iku], [ikku], [iksku], [ikstku], [ikstsku]; and [ikstski] for control) were embedded into carrier sentences like “Two keaks cookeek”. These nonsense sentences had the same syntactic structure (Det-Noun-Verb). Each target sequence was repeated 10 to 12 times at least. The presence of intervocalic consonants [k, t, s], in spite of being phonologically neutral with respect to rounding, could induce labial motion, due to their coupling with the jaw, especially for coronals (a question addressed, e.g., by Perkell and Matthies, 1992). So, we took the precaution of duplicating the whole corpus with a 4-mm bite block clenched between the left molars, in order to address the possibility of consonantal jaw interferences on the rounding pattern.

Prior to the experiment, each participant was trained to produce the sequences as naturally as possible, with a constant rate and intonation pattern. They were also instructed not to produce a silent pause between the noun and the verb in order to minimize word boundary effects on the timing of the rounding movement. Subjects received a compensation for their participation at the end of the experiment.

2.2. Audiovisual and kinematic data collection

Speakers were audiovisually recorded in the *Motor Control Lab* (McGill University, Montreal) with the Panasonic AG-DVC30 video camera available in the *Phonetics Laboratory* (UQAM, Montreal). Images were digitized at a 30Hz rate. Since one video image corresponds to two interleaved frames, we could obtain a 60Hz rate through line interpolation, synchronized with the acoustic signal recorded via a SHURE SM-86 microphone and digitized at 48KHz. The acquisition of labial shapes was performed with the ICP-Lip-Shape-Tracker. The speakers’ lips were made up with a waterproof blue lipstick for the accurate acquisition of labial contours via a numerical Chromakey processing. Front lip shape was automatically detected for each sequence with Tacle software developed at ICP.

Upper lip protrusion was simultaneously recorded via the Optotrak available at the *Motor Control Lab*. This flesh-point tracking system enabled us to capture the three-dimensional displacement of small infrared light-emitting diodes (IREDS). For the experiment, three diodes were attached with doubled-sided tape to the midline of the upper and lower lip, close to the vermilion border and to the midline of the chin. Attention was given not to obstruct inner labial contours with the infrared diodes threads. In addition, a plexiglass frame with 4 IREDS diodes was designed for references measures and attached to the speaker’s goggles. The orientation of the occlusal plane was measured by asking the subject to bite on a plexiglass frame to which three IREDS were attached. IREDS’ signals were sampled at 175Hz. A cross-correlation function (Matlab 7) was used to resynchronize the 48KHz acoustic signal from the camera with the low 10KHz signal synchronized with the Optotrak (Matlab 7). Matlab algorithms were used to rotate the data within the speaker's occlusal plane and to correct the signals for head movement.

Being aware of the high sensitivity of the Optotrak system, we only included in our analysis protrusion signals with 2mm minimum amplitude. We took this precaution to avoid monitoring any "tremor" signals instead of rounding gestures reflecting vocalic rounding anticipation control.

2.3. Articulatory analysis

For the description of rounding anticipatory behaviour, we identified the same acoustic and kinematic events and phases as previously performed with French adults (Abry *et al.*, 1996).

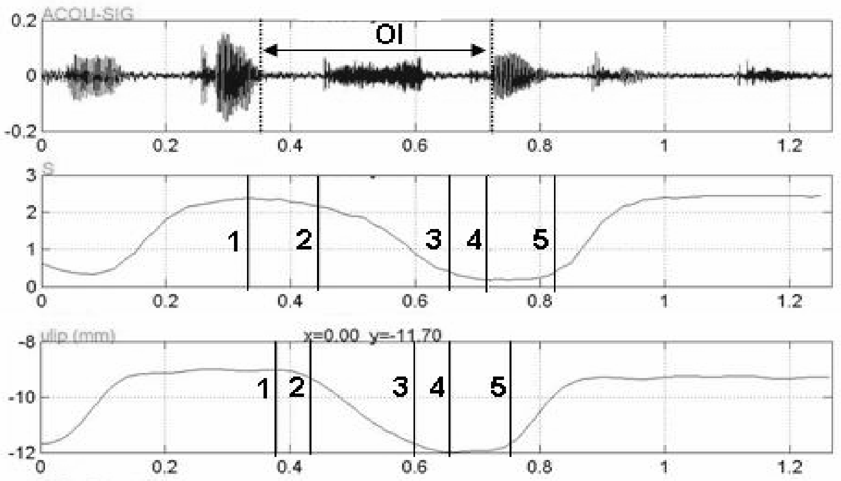


Figure 1. Sound wave (upper signal, in seconds), lip area time course (middle signal, in cm^2) and upper lip protrusion time course (bottom signal, in mm; inverted for visual correspondence with area) for a sequence [iksku] in a carrier phrase "Two keaks cookeek" uttered by a Canadian speaker (LUC) in the bite block condition. The *obstruence interval* (OI) is determined on the spectrogram by the acoustic offset of [i] and the acoustic onset of [u] formants. Identification of events on the lip area time course: 1= maximal area for [i] ; 2=10% of area difference between [i] and [u] ; 3=90% of area range between [i] and [u] ; 4= minimal area for [u] ; 5=10% of lip area difference between [i] and [u] following minimal area of [u]. Interval 2-3 corresponds to *Time Falling* (TF) and interval 3-5 to the duration of the *Hold* phase (H). Identification of events on the upper lip protrusion time course : 1= minimal protrusion for [i] ; 2=10% of protrusion difference between [i] and [u] ; 3=90% of protrusion range between [i] and [u] ; 4= maximal protrusion for [u] ; 5=10% of protrusion difference between [i] and [u] following maximal protrusion of vowel [u]. Interval 2-3 also corresponds conventionally to the duration of protrusion TF and interval 3-5 to protrusion H.

On the time course of lip area (Figure 1, middle signal), we detected maximum and minimum area values (events 1 and 4) that are both characterized by a zero value for velocity. The obstruence interval (OI) has been labelled on the acoustic signal, corresponding to the duration between the acoustic offset of the vowel [i] (characterized by the disappearance of the upper formant structure) and onset of [u] (by formants appearing again). Other robust landmarks have been selected using the classical 10% and 90% levels describing *Time Falling* and *Time Rising* in the literature on hysteresis curves. We identified the 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 [u] (event 5). The interval between event 2 and event 3 was defined as a *Time Falling* (TF) phase and the interval between event 3 and 5 as a *Hold* (H), corresponding to a phase in which acoustic efficiency of constriction area is ascertained.

On the protrusion time course (Figure 1, bottom signal; inverted for comparison sake with lip area), we also detected minimum and maximum, coinciding with a zero velocity (events 1 and 4). We selected the time corresponding to a 10% range between [i] and [u] (event 2), time corresponding to 90% of this range (event 3), and finally time corresponding to a 10% range following maximal protrusion of the vowel [u] (event 5). We used the same labels for protrusion phases (TF and H) as for area.

3. Results

3.1. Upper lip protrusion analysis

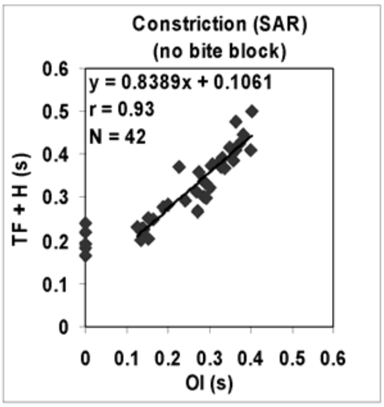
For the analysis of rounding anticipation timing, we selected the two main phases, *Time falling* (TF) and *Hold* (H) that best account for the time course of the rounding movement (without leading to a part-whole artifact when correlated with OI, Benoît, 1986). The total duration of these two phases (TF + H) has been plotted against OI (see Figures 3 and Table 1). Every speaker uttered [iu] with no interruption between the two vowels, *i.e.* with a zero OI value. Linear regression analysis (correlation coefficient and slope) was performed on data from [iCu] to [iCCCCCu], excluding [iu] sequences.

For the 4 Canadian French participants we obtained significant correlation coefficients ($p<0.0001$) and speaker-specific slopes. These slopes ranged from 0.24 to 0.65 for the no-bite-block condition and from 0.18 to 0.78 for the bite-block condition. Notice that only one subject LUC had a protrusion profile very close to a *Time-locked*, with data fairly grouped along her regression line, but sufficiently far from a zero *Time-locked* theoretical slope which would render impossible to reject the Null Hypothesis. Anyway, we will see below that this model does not correspond to her global rounding behaviour, as evidenced by the high slope of her *Constriction MEM*. The slopes as a whole did not differ from one condition to the other, except for speaker JER, with a 0.45 slope for the no-bite-block condition and a 0.78 in the bite-block condition. This can be attributed to the fact that this Canadian speaker generally displayed small protrusion amplitude so that the number of his signals decreased in the bite-block condition.

As for the English speakers, note that 3/4 participants (KAL, LAU, SAR) displayed a too small protrusion to be reliably processed in the bite-block condition. Indeed, for each speaker we could only collect about 20 signals out of 60 with at least 2mm amplitude (Table 1). In the no-bite-block condition, we could sometimes collect a greater number of relevant signals but protrusion did not generally exceed 3mm amplitude. This phenomenon has been reported for American English by Perkell and Matthies for half of their subjects (1992, p. 2914). In consequence, for these speakers, we decided to mainly rely on constriction which is an essential geometric parameter in regard to vowel acoustics. Furthermore, previous studies supported that constriction can undoubtedly account for rounding anticipation in a non protruding French adult (Abry *et al.*, 1996) and in young French children showing no protrusion yet (Noiray *et al.*, 2004).

3.2. Constriction (lip area) analysis

The total duration of constriction movement (TF+H) was again plotted against the obstruence interval (OI) (see Figure 2 for an example). Linear regression analysis performed on every [iCu] to [iCCCCCu] sequences highlighted significant correlation coefficients (p<0.0001) and speaker-specific slopes for the 8 speakers (Table 1, Figure 3). For the English participants, slopes ranged from 0.84 to 1.12 in the no-bite-block condition and from 0.75 to 1.08 for the bite-block condition. Significant correlation coefficients (p<0.0001) were also obtained for the Canadian French participants with idiosyncratic anticipatory profiles giving slopes ranging from 0.86 to 0.98 for the no-bite-block condition and from 0.85 to 1.05 for the bite-block condition. Like for the upper lip protrusion results, constriction slopes were rather close between the two conditions. Note that 2 speakers (LAU and ANN) showed a *Look-ahead* profile, with a slope near 1. This type of behaviour can be described in the framework of the speaker-specific *MEM* model, being simply a possible case, and not a general modelling. Finally, as concern the 3 non protruding English speakers, a major result is that constriction parameter lawfully accounts for their anticipatory behaviour. As a whole, these results corroborate those previously obtained for French adults and young children with the same experimental paradigm.



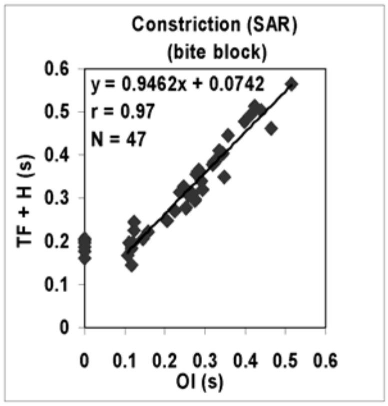


Figure 2. Rounding anticipatory behaviour for an American English speaker (SAR) in bite-block (right) and no-bite-block (left) conditions. Relation between total duration of lip constriction movement (TF+H) and duration of OI (obstruence interval), from 0 up to 4 intervocalic consonants. Correlation coefficients significant at p<0.00001.

4. Conclusion

When looking at the literature on vocalic anticipatory coarticulation over the past decades, we have come to the conclusion that it is still a tricky challenge (i) first, to accurately track labial kinematics, (ii) then, to select relevant events and landmarks for the appropriate modelisation of anticipatory strategies. In this study, we attempted to validate an experimental paradigm, extending a speaker-adaptable model – the *Movement Expansion Model* that proved to be robust for French adults and children – to speakers of English. The present results evidenced that our *MEM* model can account for the anticipatory behaviour of American English speakers as well as Canadian French speakers. Each speaker displayed his own idiosyncratic rounding movement, adopting an expansion rate with a specific slope of linear expansion around which data were *lawfully* grouped.

Another interesting finding is that upper lip protrusion cannot be generalized to the description of rounding control in every speaker of every language community. The data of our 3 non protruding English speakers corroborated this poor reliability. This phenomenon, also echoed by Perkell and Matthies (1992) can be attributed to the fact that protrusion is not in English as crucial as in French or other languages (Swedish, Greman etc.) for the rounding vowel contrast. Note that we can also encounter some French (as evidenced in Abry *et al.*, 1996, for 'Christophe') or Canadian French speakers (in a lesser extent our speaker JER) with a small protrusion. Consequently, we should focus on constriction which enables us to account for anticipatory rounding acoustic control for every speaker, even those showing a too small protrusion movement, and that is now evidenced for both French and English sound systems.

5. References

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

Abry, C., Lallouache, M.-T., and 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, 150, pages 247-255. Springer-Verlag, Berlin, Heidelberg, Tokyo, 1996.

Bell-Berti, F. and Harris K. S. Temporal patterns of coarticulation: Lip rounding. *The Journal of the Acoustical Society of America,* 71:449-459, 1982.

Benguerel, A.P., and Cowan, H.A. Coarticulation of upper lip protrusion in French. *Phonetica*, 30:41-55, 1974.

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

Henke, W. L. Dynamic articulatory model of speech production using computer simulation. Ph. D. Dissertation MIT, Cambridge, 1966.

Lubker, J. Temporal Aspects of Speech Production: Anticipatory Labial Coarticulation. *Phonetica*, 38:51-65, 1981.

Noiray, A., Ménard, L., Cathiard, M-A., Abry, C., and Savariaux, C. The development of anticipatory labial coarticulation in French: A pioneering study. In *Proceedings of INTERSPEECH, 8th ICSLP*, pages 53-56, 2004.

Perkell, J.S., and Matthies, L. M. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. *The Journal of the Acoustical Society of America*, 91(5):2911-2925, 1992.

Rubin, P., Saltzman, E., Goldstein, L., McGowan, G.R., Tiede, M., and Browman, C. Casy and extension to the task dynamic model. In *Proceedings of 4th Speech Production Seminar*, pages 125-128, 1996.

Acknowledgements. This study was supported by the Quebec's Government (FQRSC and SSHRC grants). We are grateful to David Ostry, who kindly let us perform the recordings in the *Motor Control Lab* (McGill University, Montreal) with Optotrak system. We are grateful to Johanna-Pascale Roy for her technical assistance. Finally, we thank all the participants for their interest and patience during the recordings.

Regression slopes and correlation coefficients (p<0.0001)				
	Upper lip protrusion		Constriction (lip area)	
American English subjects	Bite block condition	No bite block condition	Bite block condition	No bite block condition
DAN	0.40 (0.57)	0.51 (0.67)	0.75 (0.84)	0.87 (0.88)
KAL	Non protruding speakers		0.90 (0.82)	0.99 (0.94)
LAU			1.08 (0.78)	1.12 (0.72)
SAR			0.95 (0.97)	0.84 (0.93)
Canadian French subjects				
ANN	0.68 (0.85)	0.60 (0.92)	1.05 (0.95)	0.98 (0.99)
ANT	0.63 (0.93)	0.65 (0.92)	0.85 (0.93)	0.86 (0.94)
JER	0.78 (0.88)	0.45 (0.65)	0.94 (0.96)	0.92 (0.95)
LUC	0.18 (0.69)	0.24 (0.75)	0.86 (0.95)	0.88 (0.98)

Table 1. Regression slopes (and correlation coefficients) obtained for American English and Canadian French speakers for upper lip protrusion and constriction data (bite-block and no-bite-block conditions).

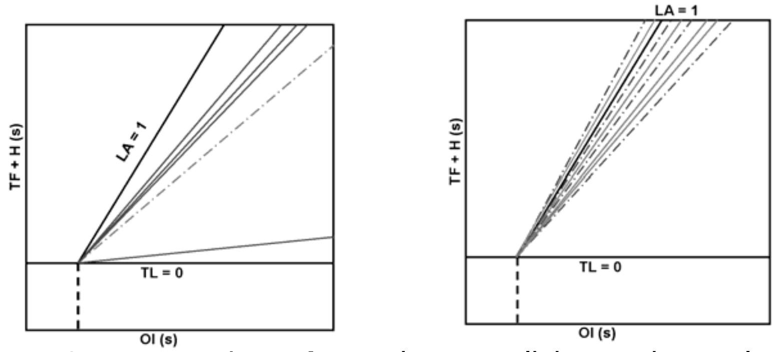


Figure 3. An overview of American English speakers' slopes (broken lines) and Canadian French (solid lines) for protrusion (left) and constriction (right) in the bite-block condition.