

Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers

Aude Noiray^{a)}

Haskins Laboratories, 300 George Street, Suite 900, New Haven, Connecticut 06511

Marie-Agnès Cathiard

Centre de Recherche sur l'Imaginaire, Université Stendhal, EA610, 38040 Grenoble cedex 9, France

Lucie Ménard

Center for Research on Language, Mind, and Brain, Département de linguistique, Université du Québec à Montréal, Case postale 8888, Montréal (Québec) H3C 3P8, Canada

Christian Abry

10 chemin du Ruy, F-38690 Chabons, France

(Received 15 January 2010; revised 13 October 2010; accepted 21 October 2010)

The modeling of anticipatory coarticulation has been the subject of longstanding debates for more than 40 yr. Empirical investigations in the articulatory domain have converged toward two extreme modeling approaches: a maximal anticipation behavior (*Look-ahead model*) or a fixed pattern (*Time-locked model*). However, empirical support for any of these models has been hardly conclusive, both within and across languages. The present study tested the temporal organization of vocalic anticipatory coarticulation of the rounding feature from [i] to [u] transitions for adult speakers of American English and Canadian French. Articulatory data were synchronously recorded using an *Optotrak* for lip protrusion and a dedicated *Lip-Shape-Tracking-System* for lip constriction. Results show that (i) protrusion is an inconsistent parameter for tracking anticipatory rounding gestures across individuals, more specifically in English; (ii) labial constriction (between-lip area) is a more reliable correlate, allowing for the description of vocalic rounding in both languages; (iii) when tested on the constriction component, speakers show a *lawful* anticipatory behavior expanding linearly as the intervocalic consonant interval increases from 0 to 5 consonants. The *Movement Expansion Model* from Abry and Lallouache [(1995a) *Bul. de la Comm. Parlée* 3, 85–99; (1995b) *Proceedings of ICPHS* 4, 152–155.] predicted such a regular behavior, i.e., a *lawful variability with a speaker-specific expansion rate*, which is *not language-specific*. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3518452]

PACS number(s): 43.70.Bk, 43.70.Mn [MAH]

Pages: 340–349

I. INTRODUCTION

Coarticulation in speech is central to the control of the articulators of the vocal tract, allowing fluency in the speech flow. For more than 40 yr, anticipatory coarticulation modeling has been greatly controversial and concurrent models have been developed to account for anticipatory behavior in different languages. Regarding the most accessible phonological feature of rounding, three main anticipation models have specifically addressed the temporal organization of the lip protrusion parameter in the vowel-to-vowel gesture from [i] to [u] or [y], separated by a variable consonantal interval.

The *Look-ahead (LA)* model predicts a protrusion movement expansion proportional to the consonant interval, starting at the acoustic offset of the unrounded vowel [i] (Henke, 1966 for English, Benguerel and Cowan, 1974 for French, and Lubker, 1981 for Swedish). The *Time-locked (TL) Model* instead posits for a temporally invariant anticipation movement: According to Bell-Berti and Harris (1982), and for a specific speech style, the rounding movement would start at

a fixed time before the acoustic onset of the rounded vowel, regardless of the length of the consonantal interval between the two vowels. The lack of empirical support for any of the classical LA and TL models led to a third model, the *Hybrid (H)* model by Perkell and Chiang (1986). In this model, the protrusion movement is divided into two steps: an initial slow phase starting at the offset of the unrounded vowel, as predicted in the LA model, followed by a faster phase, starting at a TL acceleration peak. Within the *Hybrid* framework, the more the intervocalic consonant string expands in duration, the longer the initial slow phase, while preserving a rather invariant last phase.

These three models have been tested in a major study by Perkell and Matthies (1992; the predictions of the models are presented in their Figure 1, p. 2912 and Figure 5, p. 2917). Upper-lip protrusion was measured in four adult speakers of American English. The rounding data of three out of the four participants were rather scattered and could therefore not be accounted by any of the three classical models. Only one participant showed a significant correlation coefficient (subject 2, their Figure 9, p. 2921). However, the 0.6 correlation value could be due to a part-whole correlation artifact (as demonstrated earlier by Benoît, 1986), since the two variables

^{a)}Author to whom correspondence should be addressed. Electronic mail: aude.noiray@haskins.yale.edu

“Consonant Duration” and “Onset Interval” had a common endpoint (see their Figure 3, p. 2915 and Figure 9). Finally, in line with their previous study (Perkell, 1990), the authors’ conclusion was that “data allow us to reject strong versions of all three models” (p. 280).

This statistically null result led to the development of an alternative proposal radically different from the “tug-of-war” between TL and LA concepts. The *movement expansion model* (MEM) was first developed empirically to account for anticipatory rounding behavior in French (Abry and Lallouache, 1995a,b; Farnetani, 1999, for a short presentation and concordant Italian data). The results of an experimental study on four French adults showed that the protrusion movement time (MT classically measured as the interval between “MBeg[inning]” and “MEnd” in Perkell and Matthies, 1992, their Figure 3) in relation to the duration of the consonantal obstruence interval (OI; “Consonant Duration” in Perkell and Matthies, 1992) was characterized by: (i) a rather incompressible duration for the [iy] “basic gesture,” with an execution constant; (ii) a quasi similar duration with a single intervening consonant [iCy]; (iii) a linear expansion of MT with the increase of OI from [iCy] to [iCCCCy]; (iv) with significant correlation coefficients; and (v) speaker-specific slopes for the regression lines. In this framework, provided that, for each speaker, there is a lawful (statistically significant) behavior, it remains possible that a speaker-specific slope could present a LA-like profile or a TL-profile (provided the latter is not a null hypothesis case, with a zero slope value). It should be emphasized that these cases are just possible cases, but not at all a general lawful behavior. An illustration of the MEM’s predictions for protrusion and constriction time course is provided in Fig. 1. This modeling accounts for the most salient differences between Abry and Lallouache’s studies and Perkell and Matthies’ results. Note that these French studies evidenced a regular behavior while using essentially the same measurements as for English. MT simply avoided the above-mentioned part-whole correlation; and the nearly constant MT duration between [iy] and [iCy] was accounted by computing the regression line only from [iCy] up to [iCCCCy].

In extending this protrusion model to constriction, i.e., lip area, Abry and Lallouache (1995b) determined landmarks more robust than kinematic events commonly used (velocity, acceleration), defining the two phases *Time falling* (TF) and *Hold* (H) (see below Sec. II D). In the subsequent protrusion and constriction studies, a similar procedure was adopted for protrusion. The main conceptual differences in modeling anticipation are the following: (i) As schematized in Fig. 1, the protrusion and constriction onsets (indicated by a vertical arrow) can start more or less into the unrounded vowel [i] (more for [iy] than for [iCy]), but there is essentially not any systematic difference between these two movement patterns. Hence, the protrusion and constriction onsets can start near the end of the preceding vowel or even later, as observed for the largest consonant sequence (a point on which the MEM differs from LA’s proposal, the movement onset being not locked to the [i] vowel offset). (ii) Similarly the [y] movement *onset* is not locked to the acoustic [y] onset as in TL. (iii) Notice that the maximum protrusion is not locked to [y]

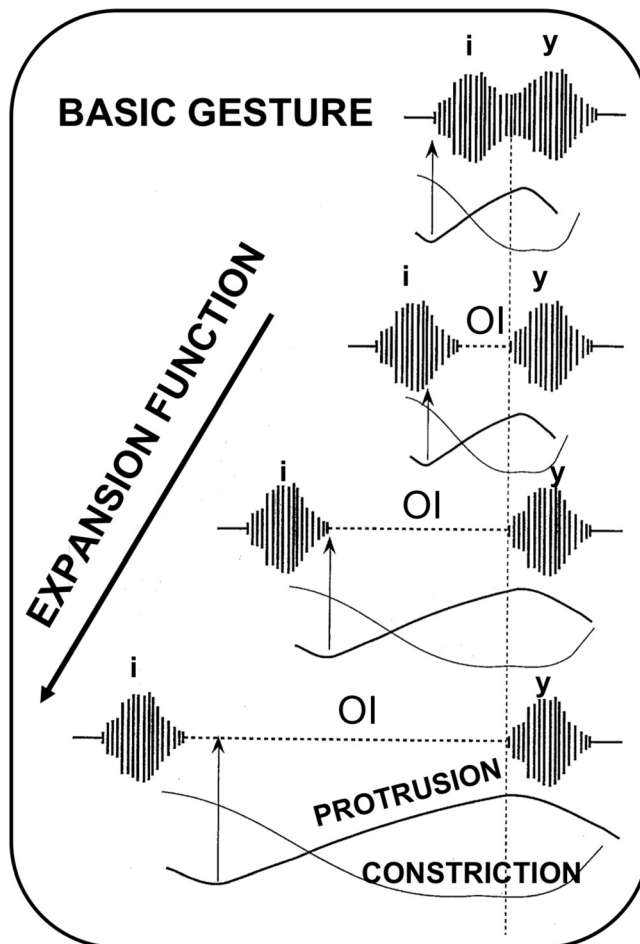


FIG. 1. MEM predictions for the time course of protrusion (bold lines) and constriction (thin lines) for the vowel [i] to the rounded vowel [y] in French. The top signal is for the *basic gesture* for rounding in the simple sequence [iy], without an intervocalic obstruent consonant (or a silent pause). The three other signals, all aligned on the [y] acoustic onset (vertical dotted line), illustrate the expansion of the rounding movement—with the prediction of a linear *expansion function*—through sequences with an increasing number of intervocalic consonants (here up to five consonants in French; the same for English [i] to [u] sequences). OI, the obstruent consonant acoustic phase between the two vowels. Vertical arrows indicate the range of variation of the protrusion and constriction onsets (see text). Notice that protrusion will be inverted for comparison’s sake in the presentation of our data.

acoustic onset, hence there is not any natural part-whole correlation. (iv) Regarding the model’s predictions, the expansion function, as already mentioned, is linear and expands at a speaker-specific rate, starting from [iCy] data and not from the y-axis intercept (as in Perkell and Matthies, 1992, Figure 5, p. 2917). (v) A lawful speaker-specific linear expansion was predicted with a statistically significant slope which may exceed 1 (i.e., LA-like with no part-whole correlation limit) and exceed 0 (i.e., above a statistically null TL). LA or TL trends can therefore be observed, as just possible speaker-specific cases, either for their protrusion and/or constriction (cf. Abry *et al.*, 1996, Figures 4 and 5, p. 254, for a presentation of the results for four French speakers, with slopes ranging from 0.42 to 0.93 for protrusion, and from 0.69 to 0.93 for constriction). (vi) Regarding a possible hybrid-like behavior for English, which is not found in French, the two sub-phases (Tf and H) were kept. Note that this is not a MEM

prediction, for which these sub-phases are not separately predicted as lawfully related to anticipation expansion. The rationale for such an output-oriented control is simply that the constricted phase (H) for the rounded vowel goal can occur any time *before* the acoustic excitation of the appropriate vocal-tract configuration by the laryngeal pulses.

Since it has been repeatedly evidenced that French and English differ in their protrusion behavior, the scattered variability observed in the study of [Perkell and Matthies \(1992\)](#) vs the *lawful variability* in [Abry and Lallouache \(1995a,b\)](#) could lead to the simple statement that there is again no winner among anticipation models (which are not limited of course to the most famous cited above for the rounding gesture). This recently led to the proposal of a compromise such as a TL model for English vs a LA model for French, Swedish, etc. Formulated by [Byrd and Saltzman \(2003, note 4, p. 157\)](#), in the context where they cited the MEM, the obvious discrepancy between results obtained for French and American could be handled in their *Task Dynamic* model with an ad hoc “side constraint,” as suggested by [Rubin et al. \(1996\)](#).

Since such a compromise did not seem satisfactory, the present study aimed at testing the MEM in both English (American) and French (Canadian) speakers. In addition to measuring upper-lip protrusion in the two languages for comparison with previous studies, we measured *constriction* that could not be investigated with flesh points techniques (cf. [Perkell and Matthies, 1992](#)) until the development of a *Lip-Shape-Tracking-System* ([Lallouache, 1991](#)). Hence, both *protrusion MEM* and *constriction MEM* were tested with English speakers, as previously with French. To minimize the influence of the mandible on protrusion movements (mainly for the lower lip) and consequently for between-lip area (constriction), we also designed a bite-block (BB) control condition.

To summarize, this study addressed the following main issues. Is anticipatory labial coarticulation a language-specific control parameter? Or could it be that one particular dimension of rounding, namely the protrusion measurement is not consistent enough to be a reliable estimate of the rounding gesture, particularly in English? Instead, could the constriction gesture be the most robust to achieve rounding and evidence a general anticipatory control pattern?

II. METHOD

A. Participants

Four adult speakers of American English (two males: MA001, MA002; and two females: FA001, FA002; ranging in age from 20 to 24, mean: 22 yr) and four speakers of Canadian French (two males: MF001, MF002; and two females: FF001, FF002; ranging in age from 22 to 35, mean: 28 yr) were recruited in the study for which they gave formal consent. American English participants (native Americans studying temporarily in Montreal) were recruited among students of McGill University, and Canadian French at UQAM in Montreal. None reported any history of hearing deficit or motor disorder. Participants were compensated for their participation in the study.

B. Stimuli

Stimuli materials were designed to investigate rounding anticipation gestures through high vowels, from front unrounded [i] to *back* rounded [u]. Since a rounded equivalent to /i/ exists in the front dimension in French but not in English, [y] was replaced with the back rounded vowel [u]. However, note that although the phonological description of [u] is similar in the two languages, their *target* configurations are different (concerning these high vowels, see [Linker, 1982](#), for their articulatory shapes; and recently [MacLeod et al., 2009](#), for their language-specific acoustic outcomes). Also, the between-language difference in *temporal organization* is reputedly a crucial contrast since, as seen above, alternative anticipation models have been supported for the two languages.

Each participant repeated a series of [iC_nu] sequences in which C_n varied from one to five consonants. The design allowed for testing an increasing consonantal interval occurring naturally in both languages. Labial consonants with intrinsic lip behavior were discarded.

Consonants [k], [s], [t] were appropriate to provide extended clusters that are legal in the phonologies of the two languages, starting from the longest [kstsk], then, by removing one consonant at once: [kstk], [ksk], [kk], and [k]. The resulting sequences—[iu], [iku], [ikku], [iksku], [ikstku], [ikstsku]—were embedded into carrier nonsense sentences, with alternating round-to-unround-to-round vowels, such as “Deux kixes coukiquent” (French), “Two keaks cookeek” (English). The complete stimuli material is provided in Appendix. Each sentence was repeated ten times in random order. Although intervocalic consonants [t] and [s] are assumed phonologically neutral with respect to rounding, the tongue-jaw coordination observed for coronals could cause labial motion as a result of their coupling with the mandible (a question addressed, but left unsolved, in [Perkell and Matthies, 1992](#)). To avoid possible consonantal jaw interferences on rounding patterns, stimuli were collected in a second experimental condition, with a 4-mm BB clenched between left molars. Each participant first performed the no-bite-blocked (NBB) condition followed by the BB one. The recording session lasted about 45 min.

C. Experimental procedure

An audio-visual recording was combined with the optoelectronic measurement of lip movements (Optotrak) at the *Motor Control Lab* (McGill University, Montreal). Prior to the recording, participants were given sufficient time to familiarize themselves with the sequences tested as well as with the BB. During the recording, participants were instructed to produce sequences as naturally as possible, with a constant rate and intonation pattern. They were also asked not to produce any silent pause between the noun and the verb in order to minimize word boundary effects on the timing of the rounding movement. For each participant, elicitation utterances were randomly prompted with a laptop monitor. Participants were comfortably seated, their head resting on the back of an armchair, facing the laptop monitor.

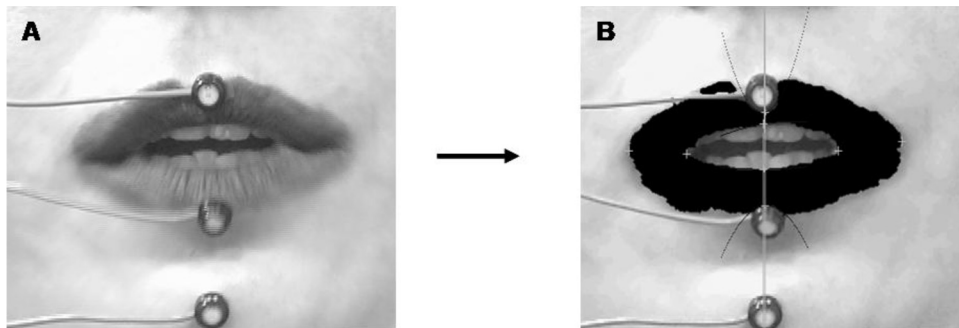


FIG. 2. Illustration of the method employed to capture upper-lip protrusion (Optotrak IREDS on the vermillion boarder of the lower and upper lips) and labial area (constriction) via a blue lip shape tracking system.

An experimenter supervised the prompting of sequences so that participants would not move.

The video camera (Panasonic AG-DVC30) was positioned in front of participants to obtain a close view of their lips. Acquisition of labial shapes was performed with the *Lip-Shape-Tracking-System* (Lallouache, 1991). Participants' lips were made up with waterproof lipstick, a "deep blue" recipe. In the frame of RGB color-coding, since the blue pigmentation is not naturally present on the face, its application on lips allowed to obtain accurate labial contours for their on-line processing via numerical Chromakey. The acoustic signal was recorded via a microphone (SHURE SM-86), pre-amplified before being recorded on the acquisition card of the camera (Canopus ADVC-100) at a 48-kHz sampling rate. Images were digitized at a 30 Hz rate (National Television System Committee norm). Since one video image corresponded to two interleaved fields, a 60 Hz rate was obtained through line interpolation. The acoustic signal from the camera was synchronous with the video data and therefore used for subsequent phonetic labeling and for labial constriction analysis.

Upper-lip protrusion was simultaneously recorded via a flesh-point-parameter tracking system (Optotrak, Certus 3020). This system allowed capturing the three-dimensional displacement of small infrared light-emitting diodes (IREDS). In this study, three IREDS were positioned with doubled-sided tape on the midline of the upper and lower lips, close to the vermillion border and on the midline of the chin (Fig. 2). IRED placement was made so as not to obstruct inner labial contours with diode wires. Upper-lip protrusion was measured as the anterior–posterior displacement of the IRED placed on participants' upper lip. A plexiglass frame with four IREDS was designed for reference measures and attached to speakers' goggles. The orientation of the occlusal plane was measured by asking the participants to bite on a triangular plexiglass frame to which three IREDS were taped. IRED signals were sampled at 175 Hz. MATLAB algorithms were used for rotation of the data within the speaker's occlusal plane and head motion correction. In addition to recording via the camera, the acoustic speech signal was also recorded on the Optotrak (sampling rate: 10 kHz) synchronously with the protrusion signal. Each recording was preceded and ended with a series of sharp bursts for post synchronization of the acoustic signals from the camera and Optotrak along with the protrusion data. A cross-correlation function was then used to synchronize both acoustic signals from the camera and Optotrak with the protrusion signal.¹

D. Data analysis

Data analysis was conducted at ICP (*Institut de la Communication Parlée*, Grenoble, France). All stimuli were labeled with PRAAT software. Between-lip shape was automatically detected for each sequence with custom-designed automatic detection software. Upper-lip protrusion and between-lip area signals synchronized with the acoustic signal were analyzed with a custom-designed speech signal editor. Protrusion signals showing amplitude lower than 2 mm were excluded from analysis. This threshold decision aimed at excluding sub-movements, for example, tremor. The acoustic and kinematic events and phases identified on the various signals were similar to those previously defined to describe anticipatory behavior for French adults (Abry *et al.*, 1996). These events are presented in Fig. 3.

On the acoustic signal, the OI (in seconds) is defined as the duration between the acoustic offset of the vowel [i] (characterized by the disappearance of the upper formant structure) and onset of [u] (upper formants appearing again). The temporal interval between the two vowels varied in relation to the number of the intervocalic consonants. On the time course of lip area (Fig. 3, middle signal) as well as upper-lip protrusion (Fig. 3, bottom signal), maximum and minimum values for [i] and [u] were measured (events 1 and 4). Both corresponded to a zero value for velocity. Acceleration events were excluded from the analyses because of their instability. More robust landmarks were selected using the conventionally 10% and 90% levels describing TF and H phase. Regarding between-lip area, the time corresponding to a 10% decrease of the area amplitude (event 2) was detected as well as the time corresponding to 90% of this range (event 3), and those corresponding to a 10% increase of lip area following minimal area of the vowel [u] (event 5). The interval between events 2 and 3 was defined as a TF phase that shows a significant decrease of lip area before reaching minimal labial constriction. This phase corresponds to the movement setting phase. The interval between events 3 and 5 delimited a H phase, corresponding to a period during which the acoustic efficiency of constriction area is about its best. Identification of events and phases on the upper-lip protrusion time course was performed using similar procedure (Fig. 3, bottom signal; inverted for comparison with lip area). Minimum and maximum values for protrusion were measured at zero velocity values (events 1 and 4). The time corresponding to a 10% range between [i] and [u] (event 2), to 90% of this range (event 3), and finally to a 10% range

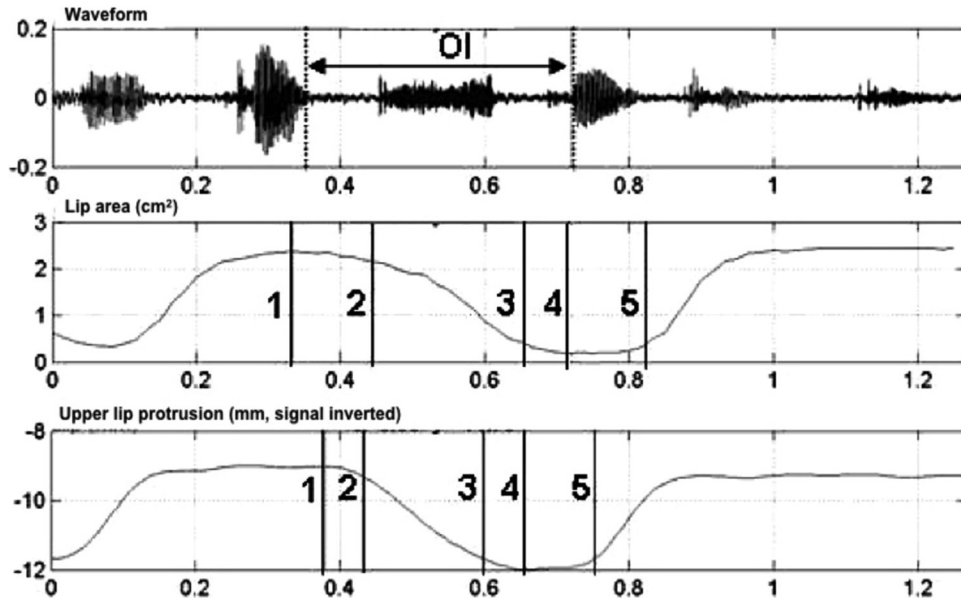


FIG. 3. Above: the waveform time (s); center: lip area time course (cm^2); below: upper-lip protrusion time course (mm; inverted for visual correspondence with constriction) for a sequence [iksku] in the sentence “Deux kixes coukiquent” uttered by a Canadian French speaker in the BB condition. The OI 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 TF and interval 3–5 to the duration of the H phase. 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.

following maximal protrusion of the vowel [u] (event 5) were selected on the protrusion signal. Similar TF and H labels as for between-lip area phases were used. Overall, these measurement events (landmarks) are more robust than movement derivatives (sometimes unstable velocity and acceleration zero crossings and peaks) leading to more stable intervals (phases).

III. RESULTS

According to the MEM parameters (Abry and Lallouache, 1995a,b), the relation between the total duration of the constriction or protrusion movement (TF + H phase in seconds) with the OI (in seconds) was preferred to evidence vowel anticipatory behavior since the variability of the two

sub-phases (TF and H) did not show any overall lawful behavior compared to [TF + H]. Presently, regarding the protrusion component, the lawfulness of the two sub-phases TF and H could hardly be tested as three American participants did not have sufficient lip movement for this measure to be useful. Regarding the constriction behavior, only three participants (two French and one American) out of eight displayed significant correlations for TF and H separately regardless of the jaw condition. Two participants (one French, one American) were “regular” for TF but not for H; one (American) for H only; two (one French, one American) did not have any significant correlation for H in BB condition. By contrast, when both TF + H were considered, all participants show a lawful (statistically significant) expansion behavior regardless of their language and the jaw

TABLE I. Regression slopes and correlation coefficients (in parentheses) of the four American English and four Canadian French speakers for upper-lip protrusion and constriction in BB and NBB conditions.

		Upper-lip protrusion		Constriction (lip area)	
		NBB condition	BB condition	NBB condition	BB condition
American English speakers	MA001	0.51 (0.67*)	0.40 (0.57*)	0.87 (0.88*)	0.75 (0.84*)
	MA002	Insufficient protrusion (<2 mm)		0.99 (0.94*)	0.90 (0.82*)
	FA001			1.12 (0.72*)	1.08 (0.78*)
	FA002			0.84 (0.93*)	0.95 (0.97*)
Canadian French speakers	MF001	0.65 (0.92**)	0.63 (0.93**)	0.86 (0.94**)	0.85 (0.93**)
	MF002	0.45 (0.65**)	0.78 (0.88**)	0.92 (0.95**)	0.94 (0.96**)
	FF001	0.60 (0.92**)	0.68 (0.85**)	0.98 (0.99**)	1.05 (0.95**)
	FF002	0.24 (0.75**)	0.18 (0.69**)	0.88 (0.98**)	0.86 (0.95**)

* = $p < 0.0001$.

** = $p < 0.00001$.

blocking condition. Note again that the two relevant phases did not share any common landmark with the consonantal interval (OI), thus avoiding any above-mentioned statistical artifact of a part-whole correlation (Benoît, 1986).

This major result fits well with the MEM prediction of an output-oriented control (see above) with a behavioral variability of the TF phase and *path* preceding the *goal* (H phase) of an acoustically sufficient rounding constriction. We mentioned that, modeling this variability via a “side constraint,” is just an *ad hoc* control solution (Rubin *et al.*, 1996).

In this section, the results obtained for upper-lip protrusion and constriction (labial area) are presented separately for American English and Canadian French participants. Linear regression analyses were conducted for each subject’s data from [iCu] to [iCCCCCu] sequences in both experimental conditions (BB and NBB) to obtain correlation coefficients (Bravais–Pearson) and slopes for the description of both protrusion and constriction individual expansion movement profiles. Since [iu] sequences were uttered without any signal interruption between the two vowels, i.e., with a zero OI value, these sequences were excluded from regression slopes. A global presentation of regression slopes together with correlation coefficients is reported for all participants in Table I.

A. Lip protrusion

From [iCu] to [iCCCCCu] sequences, different slopes were observed for the four Canadian French participants, showing a linear expansion of the protrusion movement in relation to OI duration (Fig. 4 and Table I). The slopes ranged from 0.24 to 0.65 in the NBB condition and from 0.18 to 0.78 for the BB condition, with each correlation being significant ($p < 0.00001$). One participant (FF002) displayed a behavior approximating a TL pattern. However, the regression analysis differed significantly from a zero slope (a strictly TL null hypothesis), with data well grouped along the regression lines in both NBB and BB conditions. This individual behavior can be described in the framework of the MEM, since rounding anticipation, provided that it is lawful, can display a speaker-specific expansion rate. This participant illustrates a very small movement expansion rate that was observed only for protrusion but not for constriction (cf. Sec. III B).

Student *t* test did not show any significant difference between the slopes for the BB and NBB conditions for three out of the four Canadian French participants (FF001: $t = 1.14$, $df = 93$; MF001: $t = 0.31$, $df = 88$; FF002: $t = -1.32$, $df = 96$). A significant difference was found for only one participant (MF002) who obtained a 0.45 slope in the NBB condition and 0.78 in the BB condition ($t = 2.76$, $df = 67$, $p < 0.01$). This result may be explained by a greater dispersion of his rounding data when the jaw is not constrained (NBB condition).

Three of four American English participants (MA001, FA001, FA002) displayed protrusion amplitudes that were too small (less than 2 mm) to be reliably processed, contrary to the Canadian French participants (ranging from 3.7 to 6.7

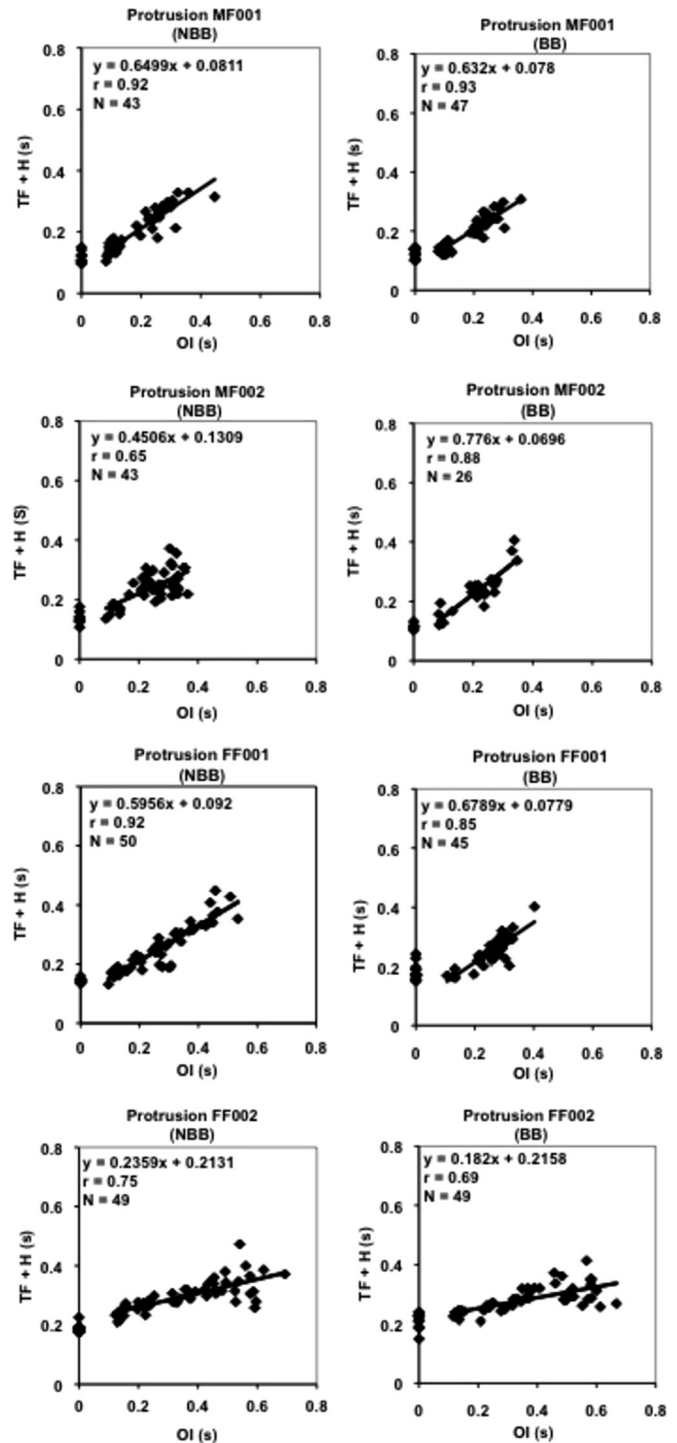


FIG. 4. Rounding anticipatory behavior for upper lip protrusion in four Canadian French speakers in bite-block (BB, right) and no-bite-block (NBB, left) conditions. Regression slopes and correlation coefficients (significant at $p < 0.00001$) between total duration of lip constriction movement [TF + H] and OI duration in seconds.

mm). Thus, although most descriptions of rounding anticipation have focused on upper-lip protrusion, the reliable use of this measure was only evident in our study for one American English participant (MA001), whose protrusion behavior (up to 6.9 mm) was equivalent to the most protruding Canadian French. This finding corroborates previous reports on the instability of protrusion parameter in English (1.89 mm protrusion reported in Perkell and Matthies, 1992, vs up to

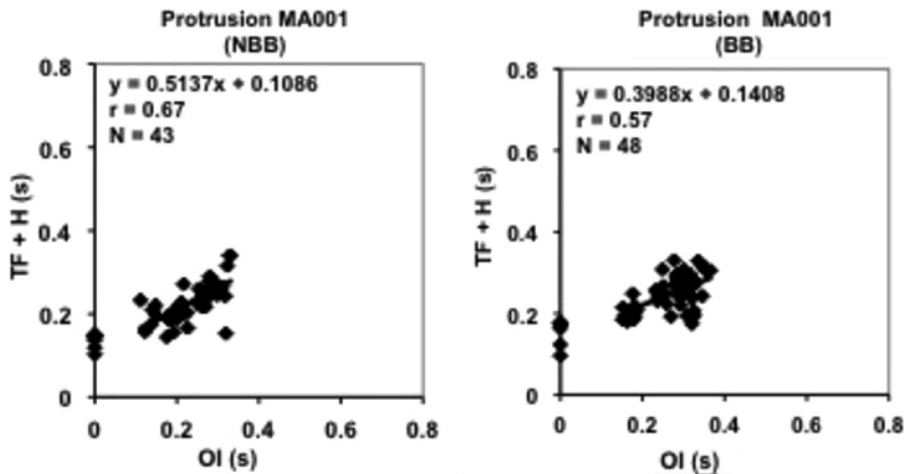


FIG. 5. Rounding anticipatory behavior for upper-lip protrusion for the American English participant who displayed a sufficient protrusion in bite-block (BB, right) and no-bite-block (NBB, left) conditions. Regression slopes and correlation coefficients (significant at $p < 0.00001$) between total duration of lip constriction movement [TF + H] and OI duration in seconds.

12 mm in Swedish, Lubker and Gay, 1982). The rounding movement of the only American English participant, whose protrusion could be considered, increased linearly with OI duration (Fig. 5), with a personal expansion slope (0.51 in NBB condition and 0.40 in the BB condition; no significant difference: $t = 0.94$, $df = 89$), thus supporting a MEM behavior.

B. Labial constriction

The time course of labial constriction was also described as the relation between the [TF + H] phases and OI duration, using the same statistical procedure. Regarding the expansion of the constriction gesture, slopes obtained for the four Canadian French speakers were relatively high, ranging from 0.86 to 0.98 in the NBB condition and from 0.85 to 1.05 in the BB condition (Fig. 6). For each participant, the correlations were statistically significant ($r > 0.90$, $p < 0.00001$) and Student t test did not show any significant difference in constriction movement between the two experimental conditions (FF001: $t = 1.20$, $df = 81$; MF001: $t = -0.19$, $df = 86$; MF002: $t = 0.36$, $df = 93$; FF002: $t = -0.25$, $df = 106$).

Among the Canadian French participants, one participant (FF001) displayed a typical LA pattern with a 0.98 slope in NBB condition and 1.05 in BB condition. As mentioned in the Introduction, this type of behavior is predicted in the framework of the speaker-specific MEM modeling as a possible individual case of lawful variability, without implying any generalization of an LA modeling to all speakers of a language (Abry and Lallouache, 1995b). Note that the participant (FF002) who displayed very low protrusion slopes (NBB: 0.24 and BB: 0.18, Fig. 4) had on the contrary constriction slopes (NBB: 0.88 and BB: 0.86) rejoining values met in the French group.

Unlike upper-lip protrusion, the analysis of constriction parameter provided vocalic rounding data for every American English participant. From [iCu] to [iCCCCCu] sequences, the four participants displayed speaker-specific rounding patterns (Fig. 7) with slopes ranging from 0.84 to 1.12 in the NBB condition and from 0.75 to 1.08 in the BB condition ($p < 0.0001$). For each subject, the anticipatory profiles did not differ significantly between the two experimental conditions (MA001: $t = -1.15$, $df = 79$; MA002:

$t = 0.74$, $df = 70$; FA001: $t = -0.18$, $df = 89$; FA002: $t = -1.55$, $df = 73$).

Figure 8 provides a summary of the protrusion (left panel) and constriction (right panel) patterns (TF + H) found for English (dotted-dashed lines) and French speakers (solid lines) in the BB condition—the one that minimizes jaw movement effects. This schematic graphic summary is an indispensable complement, proving that results summarized by the schemas in Abry *et al.* (1996, Figs. 4 and 5, p. 254, for four French speakers) have now been successfully extended for English speaking subjects toward a generalization of the output-oriented constriction MEM. On the right panel, all speakers meet in a bundle of slopes.

IV. CONCLUSIONS

The present study aimed at comparing vocalic rounding anticipation between two linguistic communities—American English and Canadian French in the time course of the vowel gesture from [i] to [u]. Rounding behaviors were tracked via measurements of upper-lip protrusion flesh points and with a specific image processing system for labial constriction on [iC_nu] sequences varying in the length of the consonantal interval. The results indicate that for a general framework to be applied to both languages (which have repeatedly been reported as different in the implementation of the rounding gesture), upper-lip protrusion is not an appropriate measure. This parameter has proven to be inadequate because many speakers have movement amplitudes that are too small for accurate measurements and this observation was also true for one of four French speakers tested in a previous study (Abry *et al.*, 1996). In this study, protrusion could only be confidently measured for one English speaker out of four. Investigating the control of the vocal-tract area at its output—namely lip constriction—provided generalized results for the two communities. Speakers displayed lawful movement expansion behaviors, with statistically significant correlations between the duration of the intervocalic consonantal interval and the time course of anticipatory rounding. Since the between-lip area parameter could be more influenced by the mandible than upper-lip protrusion, we designed two experimental conditions with and without a BB. Only one participant out of eight (an American one) obtained more scattered

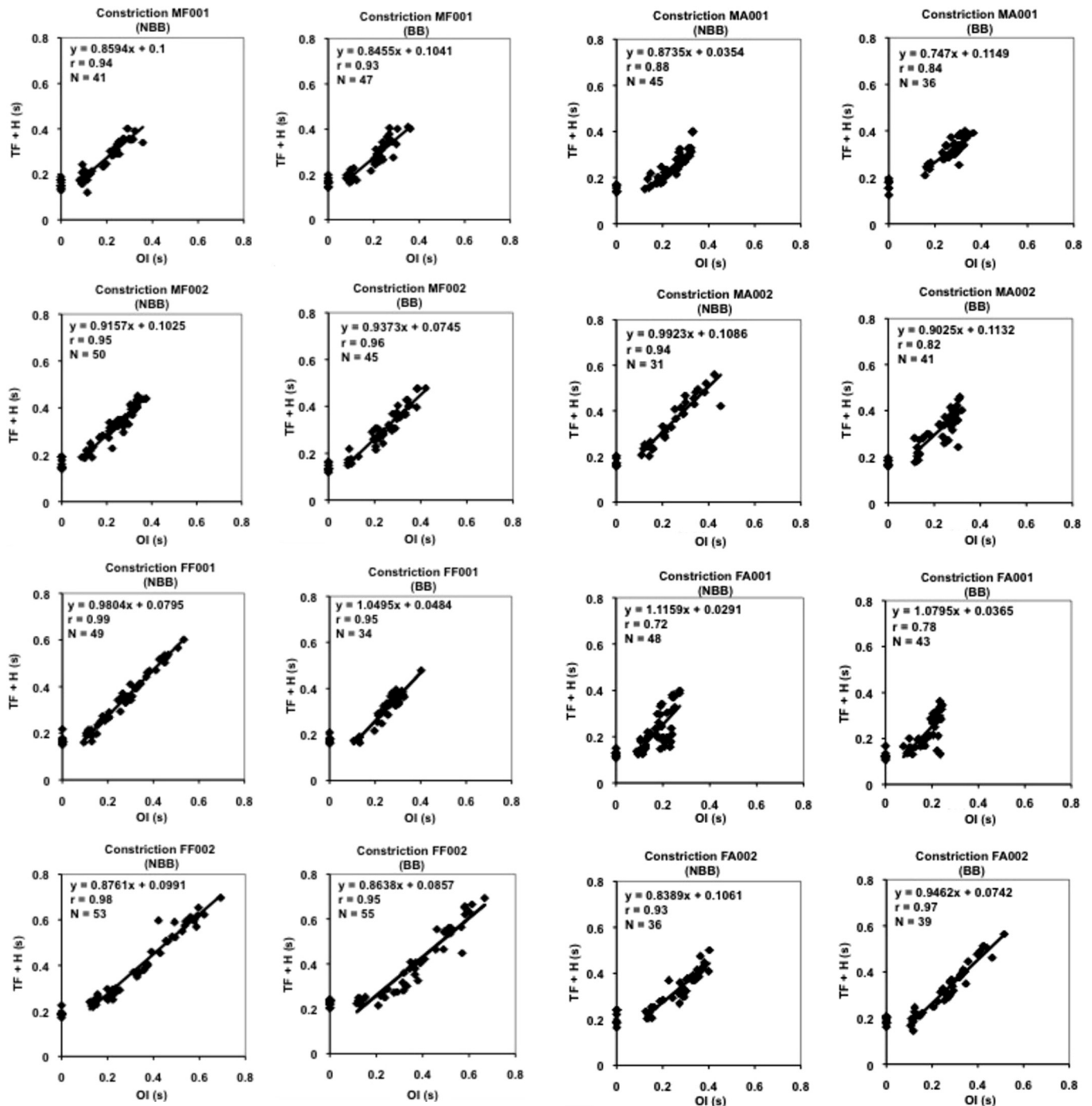


FIG. 6. Rounding anticipatory behavior for between-lip area (constriction) in four Canadian French participants in bite-block (BB, right) and no-bite-block (NBB, left) conditions. Correlations (significant at $p < 0.0001$) between total duration of lip constriction movement [TF + H] and OI duration in seconds.

results (yet with a significant correlation) without BB. Figure 8 provides an overview of the constriction profiles obtained for all speakers in the BB condition (the least influenced by the jaw). It shows that there is no any language-specific trend concerning both the statistical significance of correlation regression lines or differences in their slopes. Indeed, neither a general trend toward a TL model (observed only for the protrusion data of one French subject), nor toward a LA behavior was observed, as demonstrated by the presence of

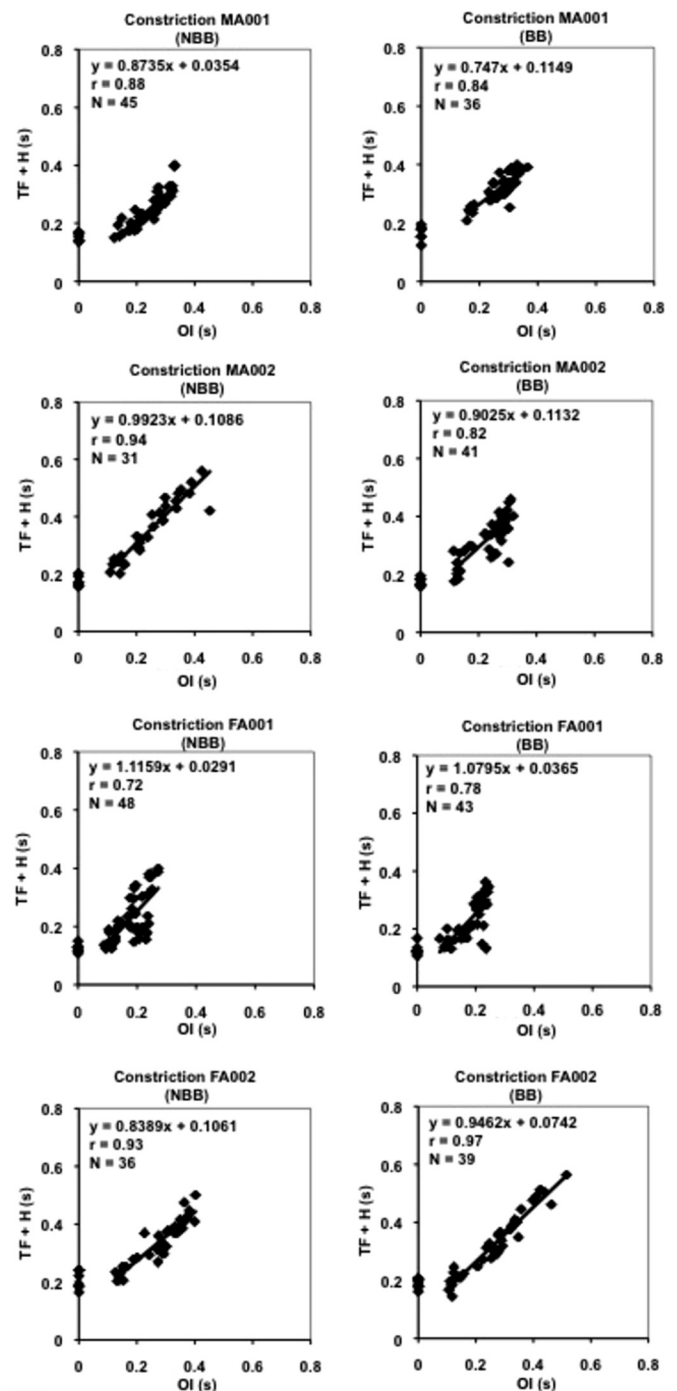


FIG. 7. Rounding anticipatory behavior for between-lip area (constriction) in four American English participants in bite-block (BB, right) and no-bite-block (NBB, left) conditions. Correlations (significant at $p < 0.0001$) between total duration of lip constriction movement [TF + H] and OI duration in seconds.

low slope values, the lowest being 0.75 (0.69 in [Abry et al., 1996](#)). But what was observed was a *speaker-specific* expansion rate, with *no grouping by language*. Analyzing TF and H phases separately for testing a possible Hybrid Model did not evidence any overall lawful behavior. But, taken together, TF + H phase account for the regular expansion control from rounding onset to rounding offset. In control terms, this means that such a behavior in which the phase, say the path (TF), that precedes the goal (H phase), can be highly variable across

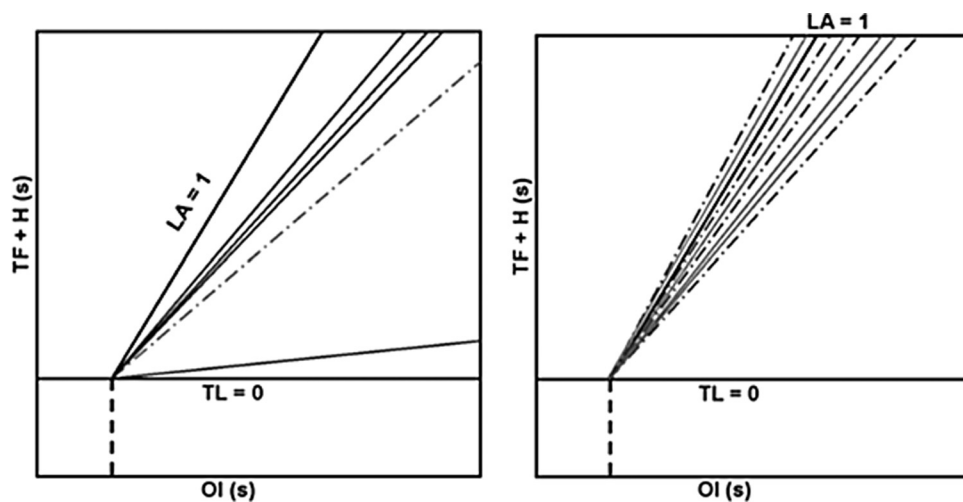


FIG. 8. An overview of expansion rates (slopes) for American English (dotted-dashed lines) and Canadian French (solid lines) for protrusion (left) and constriction (right) in the bite-block condition. (Only one English speaker had measurable data for protrusion). The slope of 1 is predicted by the LA model, while zero slope is for TL model. For simplicity's sake, slopes are shown to change from one consonant interval (vertical broken lines) to many.

speakers, must be interpreted from a general point of view as an output-oriented control. The goal constricted phase (H) for the rounded vowel, with an acoustically sufficient constriction, can thus occur any time before the acoustic excitation of the appropriate rounded vocal-tract configuration by the laryngeal pulses (Cathiard *et al.*, 1996, note 1, p. 219).

The data collected in this study support the predictions formulated within the MEM framework with idiosyncratic movement expansion rate. For most languages that use the common lip constriction gesture for the corner high back vowel [u], the core behavioral timing is related to the more efficient method of changing the area at the output of the vocal tract. It has been shown in a companion study with French children (Noiray *et al.*, 2008) that the timing pattern [TF + H] of the path toward this articulatory-acoustic goal could be acquired as soon as the age of 3.5 yr, and generally achieved by 5.5 yr. A comparable study has to be conducted with English speaking children to assess the age at which this general lawful behavior would be mastered in this community.

In summary, anticipatory labial coarticulation is not a language-specific control parameter for vowels. This is evidenced if one definitely does not rely on such an inconsistent component as protrusion, but on the timing pattern of the rounding gesture tested within the robust phases given by the lip *constriction* maneuver. Consequently there is no need for a compromise (as proposed by Byrd and Saltzman, 2003) in the good-old-fashioned “tug-of-war” between the LA and TL modeling stances. Constriction MEM is now the winner for both French and English speakers. Further work is needed to test these results more broadly with more speakers and across other languages.

ACKNOWLEDGMENTS

This study was supported by Fond Québécois de la Recherche sur la Société et la Culture (FQRSC) and Social Sciences and Humanities Research Council (SSHRC) grants (Quebec’s Government), the French Ministry of National Education and Research (Ph.D. fellowship), and NIH Grant No. DC-02717. The data were collected at Motor Control Lab (McGill University, Montreal) and the analysis performed at ICP (Institut de la Communication Parlée, Grenoble, France).

We are grateful to Jérôme Aubin and Christophe Savariaux for their experimental and data processing support and to Johanna-Pascale Roy for her technical assistance. Finally, we are grateful to all the participants for their interest and patience during the recordings. A preliminary part of this study has been presented at the *7th International Seminar on Speech Production* (Ubatuba, Brazil, 2006).

APPENDIX

TABLE II. List of stimuli used for the recording of the Canadian French speakers.

Sequences Article-noun-Verb	Transitions V1 ... C _n ... V2	Number of consonants
Deux ki oukiquent	iu	0
Deux ki coukiquent	iku	1
Deux kikes coukiquent	ikku	2
Deux kixes coukiquent	iksku	3
Deux kixtes coukiquent	ikstku	4
Deux kixtes scoukiquent	ikstsku	5
Deux kixtes skikiquent	ikstski	5

TABLE III. List of stimuli used for the recording of the American English speakers.

Sequences Article-noun-Verb	Transitions V1 ... C _n ... V2	Number of consonants
Two kea ookeek	iu	0
Two kea cookeek	iku	1
Two keak cookeek	ikku	2
Two keaks cookeek	iksku	3
Two keakst cookeek	ikstku	4
Two keakst skookeek	ikstsku	5
Two keakst skeekeek	ikstski	5

¹As the sampling rate of the acoustic signal from the camera was greater (48 kHz) than for the Optotrak (10 kHz), the camera acoustic signal was downsampled to 10 kHz to match the one recorded via the Optotrak (PRAAT function). A temporal window was selected in both signals to be compared and a cross-correlation function was computed to measure their difference in samples (MATLAB 7. Mathworks). The bursts were used in the cross-correlation function as references for the realignment of both acoustic signals.

- Abry, C., and Lallouache, T. (1995a). "Le MEM: Un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français" ("MEM: A speaker parameterized anticipation model)," *Bul. de la Comm. Parlée* 3, 85–99.
- Abry, C., and Lallouache, T. M. (1995b). "Modeling lip constriction anticipatory behaviour for rounding in French with the MEM," in *Proceedings of ICPHS*, 4, Stockholm, Sweden, pp. 152–155.
- Abry, C., Lallouache, M. T., and Cathiard, M. A. (1996). "How can coarticulation models account for speech sensitivity to audio-visual desynchronization?" in *Speechreading by Humans and Machines*, edited by D. Stork and M. Hennecke (NATO ASI Series F: Computer and Systems Sciences, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo), Vol. 150, pp. 247–255.
- Bell-Berti, F., and Harris, K. S. (1982). "Temporal patterns of coarticulation: Lip rounding," *J. Acoust. Soc. Am.* 71, 449–459.
- Benguerel, A. P., and Cowan, H. A. (1974). "Coarticulation of upper lip protrusion in French," *Phonetica* 30, 41–55.
- Benoît, C. (1986). "Note on the use of correlations in speech timing," *J. Acoust. Soc. Am.* 80(6), 1846–1849.
- Byrd, D., and Saltzman, E. (2003). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," *J. Phonetics* 31, 149–180.
- Cathiard, M.-A., Lallouache, M.-T., and Abry, C. (1996). "Does movement on the lips mean movement in the mind?" in *Speechreading by Humans and Machines*, edited by D. Stork and M. Hennecke (NATO ASI Series F: Computer and Systems Sciences, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo), Vol. 150, pp. 211–219.
- Farnetani, E. (1999). "Labial coarticulation," in *Coarticulation: Theory, Data, and Techniques*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge), pp. 144–163.
- Henke, W. L. (1966). "Dynamic articulatory model of speech production using computer simulation," Ph.D. Dissertation, MIT, Cambridge.
- Lallouache, M. T. (1991). "Un poste «Visage-parole» couleur. Acquisition et traitement automatique des contours des lèvres (A" «face-speech» inter-face. Automatic acquisition and processing of labial contours)," Ph.D. ENSERG, Grenoble, France.
- Linker, W. (1982). "Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study," *Working Papers in Phonetics, WPP* 56, 1–138.
- Lubker, J. (1981). "Temporal aspects of speech production: Anticipatory labial coarticulation," *Phonetica* 38, 51–65.
- Lubker, J., and Gay, T. (1982). "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables," *J. Acoust. Soc. Am.* 71(2), 437–448.
- MacLeod, A. A. N., Stoel-Gammon, C., and Wassink, A. B. (2009). "Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers," *J. Phonetics* 37, 374–387.
- Noiray, A., Cathiard, M. A., Ménard, L., and Abry, C. (2008). "Emergence of a vowel gesture control. Attunement of the anticipatory rounding temporal pattern in French children." in *Emergence of Language Abilities*, edited by S. Kern, F. Gayraud, and E. Marsico (Cambridge Scholars Publishing, Newcastle, UK), pp. 100–116.
- Perkell, J. S. (1990). "Testing theories of speech production: Implications of some detailed analyses of variable articulatory data," in *Speech production and speech modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht), pp. 263–288.
- Perkell, J. S., and Chiang, C. (1986). Preliminary support for a "hybrid model" of anticipatory coarticulation," in *Proceedings of 12th International Conference of Acoustics*, Toronto, pp. A3–A6.
- Perkell, J. S., and Matthies, L. M. (1992). "Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability," *J. Acoust. Soc. Am.* 91(5), 2911–2925.
- Rubin, P., Saltzman, E., Goldstein, L., McGowan G. R., Tiede, M. and Browman, C. (1996). "CASy and extension to the task dynamic model," in *Proceedings of 4th Speech Production Seminar*, pp. 125–128.