Why do syllable onsets attract consonant(s)?

Nathalie Vallée, Thi Thuy Hien Tran, Paolo Mairano, Silvain Gerber & Solange Rossato

The present study is carried out in the context of a project whose primary aim is to define the physical components that could help explain what makes a syllable. We examine whether phonetic factors can account for observed universal trends in syllable structure, particularly those that provide an explanation for the Maximum Onset Principle (MOP). MOP forms the basis of syllabification theories allocating consonantal segments to syllable-initial position rather than to syllable-final position. Two theoretical frameworks can provide a phonetic basis for MOP, namely the Frame/Content Theory and Articulatory Phonology, but they do not account for it thoroughly. Consequently, we propose to reconsider the MOP in the light of new articulatory data focusing on basic jaw movements, and assume that MOP results from asymmetric jaw oscillation cycles. Kinematic data were obtained using a Carsten 2D Electromagnetic Articulograph (AG-200 EMA). Jaw oscillations were investigated for duration, velocity, amplitude of the opening vs closing phases, and then statistically analyzed according to different syllable structures. The results on French and Vietnamese bring out differences between the opening and closing phases, showing clear duration and velocity asymmetries. These results are, however, not consistent with previous works on American-English which can provide a phonetic explanation for MOP, based on the properties of the jaw cycles.

1. Introduction

Syllables with an asymmetric structure, i.e. syllables with different onset and coda structures, are frequently found in the world's languages. For example, CV syllables (empty coda), CCVC (complex onset and simple coda), and VC (empty onset) have an asymmetric structure. We call this asymmetric syllable structure 'phonological asymmetry' henceforth. The present study attempts to provide an explanation for this cross-language tendency. Providing the link between phonetic and phonological asymmetries in the structural organization of syllables should allow an explanation of the Maximum Onset Principle (MOP; Kahn 1976; Selkirk 1982; Clements & Keyser 1983; Goldsmith 1990; Blevins 1995). This universal principle of syllabification supports the view that consonants tend to be syllable-initial rather than syllable-final; consonants tend to be attracted in the prevocalic position (thus acting as onsets) rather than in the post-

Italian Journal of Linguistics, 27.1 (2015), p. 133-160 *(received September 2014)*

vocalic position (as codas). MOP is supported by a relatively large number of empirical observations and is widely used in linguistics, psycholinguistics, or in natural language processing. From a theoretical point of view, two frameworks can provide a phonetic base of MOP: the Frame/Content Theory (MacNeilage 1998, 2008) and Articulatory Phonology (Browman & Goldstein 1988, 1990, 1995, 2000).

According to the first theoretical framework, the universal organization of speech is supposed to be a result of jaw oscillation. Indeed, the Frame/Content Theory suggests that a close link exits between the universal CV structure and the cycle of jaw movement: a consonant is produced when the jaw is in a high position whereas a vowel is made when the jaw is in a low position. The cycle of jaw movements thus provides the "Frame" while the segments (consonants and vowels) are said to determine the "Content". The Frame/Content model also predicts that CV sound combinations favored in natural languages are similar to children's productions at the stage of babbling, irrespective of the language(s) spoken in their linguistic environment (Davis & MacNeilage 1995, MacNeilage & Davis 2001). These favored combinations called Pure Frames are produced with a single cycle of jaw oscillation without front-back movement of the tongue during consonant-vowel productions. As a result, they are considered simple to produce with a minimum articulatory cost, what should explain their overrepresentation cross-linguistically. However, Rousset (2004), then confirmed by Vallée *et al*. (2009), showed that the proportion of Pure Frames among VC sequences found in the lexicon of 15 languages was significantly higher than the proportion of Pure Frames among CV syllables. This observation suggested a stronger articulatory cohesion between vowel and consonant when the latter is in syllable-final position compared to when it is syllable-initial. The authors thus interpreted such result as evidence in support of an onset vs rime syllable structure, as proposed by those phonologists who argue that syllable structure is rather hierarchical than linear (among others, Klein 1993).

An unsatisfactory feature of the Frame/Content Theory is that it does not provide any clear explanation as to why the preferred structure is consonant-vowel rather than vowel-consonant. Yet the VC structure matches as perfectly as CV with a single jaw cycle, but it represents less than 5% of the syllables found in the world's languages (Rousset 2004). Another weak point of the Frame/Content Theory which should be noted, is that the CVC structure is the second highest syllable type in the languages of the world, and the highest in some languages (for example, languages of Southeast Asia like

Nyakhur, Thai, Vietnamese, Wa, etc). This syllable type can still contain Pure Frames in it (in onset-nucleus or in nucleus-coda). However, Vallée *et al*. (2009) have found that Pure Frame CVC syllables are underrepresented in the world's languages, since consonant place repetition within a syllable (e.g. /pap/ and /tat/ syllables compared to /pat/or /tap/) are largely avoided.

According to a different perspective, Articulatory Phonology suggests that jaw has no direct role in syllable production, and is considered as merely supporting the actions of others articulators, namely, the lower lip and the tongue. The universal predominance of the CV structure is explained by natural features of consonants, which are "in-phase coupling" with the following vowel. Articulatory gestures for a C and a subsequent V are closely coordinated and compatible with each other without negative mutual interference, because they are realized thanks to independent articulatory mechanisms that allow, at the same time, a synchronized trigger of gestures for C and V. By analyzing gestural overlap, substantially estimated in the time domain by measures of the C-Center Effect (Browman & Goldstein 1988), Articulatory Phonology states that CV combination is "the most stable coupling mode", and the different intra-syllable positions which are allowed for C (onset or coda) correspond to different configuration modes of gestures (Browman & Goldstein 1995; Byrd 1995). Within a syllable, onset and coda consonants show differences in the temporal organization of articulatory gestures with the vowel. Measurements of the time interval between the consonantal target (or the mean of the consonantal targets in the case of clusters; C-Center) and the vowel target indicate a much tighter and global coordination with the vowel gesture for prevocalic consonants. For syllable-final consonants, the gestural coordination is observed between the first postvocalic-consonant and the vowel. This suggests a global syllable gesture for onsets and a more limited gestural coordination for codas. In contrast to the Frame/Content Theory, the syllable corresponds in Articulatory Phonology to "a gestural constellation", which is the product of several individual actions corresponding to coarticulated segments. The favored CV combinations are those for which gestures are most naturally in phase and therefore stable (Goldstein *et al.* 2006, Whalen *et al*. 2011) while the reverse VC structure seems to require a more local, less cohesive and less adaptable motor coordination relative to the vocalic gesture. Moreover, in coda position it is expected that the production of the consonant, when anticipated, masks the acoustic realization of the vowel. Consequently, a VC sequence shows much less stability in the coordination of gestures and shows no C-Center effect. This fact may also explain the vulnerability of consonants in syllable-final position, which are more subject to lenition than consonants in syllable onset. One question that still remains to be investigated is whether the higher proportion of Pure Frames within VC syllables compared to CV syllables can also be explained by the gestural characteristics of the nucleus-coda coordination.

A possible explanation of this apparent nucleus-coda coarticulation can be deduced from the results obtained by Kelso et al. (1985), Gracco (1994), Redford (1999) and Redford & van Donkelaar (2008). Gracco (1994) investigated the opening and closing phases of the jaw cycles in /sV_1CV_2/ sequences, with C=/p b m/, V_1=/i æ/ and V_2 =/æ/. In Redford's study, four syllable types (CV/ʃ/, /ʃ/VC, /s/CV/ ʃ/, / ʃ/ VC/s/ where V=/i a u/ and C=/p t k/) were chosen to account for differences between the two phases of the jaw cycles in closed syllables (CVC, CCVC and CVCC). From continuous repetitions of five syllable types (CV, VC, CVC, CCV and VCC), Redford & van Donkelaar (2008) attempted to demonstrate that the open-close jaw cycles are consistent with the syllable structure. They were not able, however, to clarify in what extent the jaw cycles influenced the phonetic realization of speech segments and, conversely, in what extent speech segments constrained jaw movements. Kelso and colleagues' study was designed to examine temporal organization of speech by removing as much as possible the individual characteristics of speech segments and preserving the prosodic pattern. The spatiotemporal characteristics of CV forms were analyzed from reiterant speech in which normal syllables of two sentences were replaced by /ba/ or /ma/. They found differences between the opening and closing phases of the jaw cycles at the level of duration, velocity and displacement, and linked them to the dynamic properties of the cycles.

The consistency of results across these four studies with different design, may lead to a phonetic explanation for MOP. By seeking to understand the spatiotemporal organization of speech or seeking to evaluate the impact of biomechanical properties of the jaw on the articulation of segments, these studies revealed an asymmetry between the two phases of the jaw cycle: (i) the raising phase (closing) involves a greater peak of velocity than the lowering phase (opening); (ii) the opening phase has a greater duration than the closing phase; (iii) the lowering phase is articulated with a greater displacement (distance between minimum and maximum opening positions) compared to the raising phase; (iv) the degree of jaw stiffness (estimated by calculating the slope of the correlation between displacement and velocity) is found to be smaller for the opening phase than for the closing phase. In Gracco's study, the opening and closing phases were found to covary suggesting that oral openings and oral closings are not independent events, as for instance, the covariation found between the jaw opening velocity (for the vowel) and the subsequent closing velocity (for the consonant). Furthermore, Gracco suggested that oral closing depends on oral opening, by the fact that the faster the opening the faster is the closing. A possible explanation given by Gracco is that the oral opening, associated with the vocalic sound, is generally slower because it is involved in shaping resonance and providing the correct VOT, while the closing must be faster to integrate the laryngeal and supralaryngeal combined adjustments.

Following the findings of these previous studies, we assume that if these asymmetries in the properties of the jaw cycle were confirmed in a large amount of languages, then they could be said to account for several tendencies found in the world's languages (see, among others, Rousset 2004): (i) CV rather than VC as the universal canonical syllable; (ii) strong preference for syllable-initial clusters over syllable-final clusters; (iii) complex consonants (e.g. prenasalized, labialized, clicks, affricates, doubly-articulated stops, palatalization) uncommon in coda position; (iv) larger proportion of Pure Frames in nucleus-coda sequences (rime) than in onset-nucleus sequences; (v) cross-linguistically, reduction in the size of phoneme inventories and predominance of consonant lenitions in syllable-final position (constraints on features).

The purpose of the current study is to evaluate whether the jaw cycle plays a role in the understanding of the physical nature of syllables. While previous works (cited above) focused mainly on American English, our ongoing study involves a broader multilingual perspective (French, Polish, Brazilian Portuguese, Ruwund, Tachelhit, and Vietnamese) to find out features of the jaw cycle that could provide a phonetic basis for MOP. The selected languages show various degrees of complexity in their syllable structures and consonant realizations, as well as differences in coda lenition processes and phonological restrictions on codas.

In this paper, we present the first results obtained for an investigation of French and Vietnamese speech. Although only reiterant speech productions of VV sequences are analyzed to date for Vietnamese, the reason why we opted first to compare these two languages needs clarification. Vietnamese is traditionally described as a monosyllabic tone language (Đoàn 1999). Much dissimilarity between French and Vietnamese refer to their lexical and syllabic structures (Tran, 2011). The Vietnamese syllable types are captured by the pattern C1(w)V(C2), the brackets indicating the optional constituents. This pattern implies that the glottal stop which always appears in onset position is phonemic, according to Đoàn (1999). Syllabic structure diversity is much more present in French, due to the fact that complex onsets and codas are allowed (C1)(C2)(C3)V(C4)(C5)(C6)(C7) (Rousset, 2004). The most favored syllable types in both languages are CVC and CV. But Vietnamese is predominantly a CVC language while French is mostly a CV language. Like other Southeast-Asian languages, Vietnamese is typologically an isolating CVC language. Therefore, in contrast to French, Vietnamese favors a symmetrical syllable structure and disallows complexity in onset and coda positions.

2. Methodology

2.1. Experimental setup

Tongue, jaw, and lip movements were recorded during a speech production task using the electromagnetic articulography system (EMA) AG200 developed by Carstens®. The acquisition in 2D was conducted at a frequency of 200 Hz, thanks to five coils attached to the articulators (jaw, lower lip, upper lip, tongue tip and tongue body), and two reference coils (attached on the bridge of the nose where the skin movement is minimal and above the upper central incisors) for aligning head movements on the mid-sagittal plane of the subject.

The acoustic signal was recorded via a digital Marantz PMD670 stereo recorder and a directional microphone AKG C1000S, and subsequently digitized at 44.1 kHz.

2.2. Corpus and participants

The preliminary results presented in this paper come from five repetitions of three pseudowords pronounced at a normal speech rate by French speakers:

- /aiaiaiaiai/ (5 jaw cycles)
- /tataltat/ (3 jaw cycles)
- /sasalsas/ (3 jaw cycles)

Regarding Vietnamese, the first results presented here are based on five repetitions of the /aiaiaiaiai/ sequence pronounced by Vietnamese speakers.

These data were extracted from a larger corpus containing other pseudowords and other structures: CV.LVC, CLV.CVLC, CLV.CVCL, CLV.CLVC; C=/p b t d s z/, L=/l/, V=/i a/; the point represents a syllable boundary, and a list of sentences. For Vietnamese, the pseudowords were monosyllabic sequences C_1V and C_1VC_2, with C_1=/b t d s z/, C_2=/p t/ (only /p/ and /t/ are used as coda consonants in Vietnamese), V=/i a/ and modal tone B1-D1.

Pseudowords were used in this study because they provide the possibility to avoid as much as possible the influence of both lexical knowledge (frequency or familiarity) and semantic processing. Pseudowords were also used to control phoneme distribution, and smooth any contextual effects in the production of sound sequences. We varied the vowel to explore the effect of the degree of opening on jaw cycle. The consonants are the same that those used in the previous studies cited above. We analyzed firstly alveolar obstruents /t s/ because they share the highest jaw position, and show the greatest vertical displacements of the jaw when coarticulated with low vowels.

Five female native-speakers of French (CC, EB, LV, ML, SB) aged from 20 to 43 and five female native-speakers of Vietnamese aged from 28 to 35 and coming from Northern Vietnam (HOA, LINH, THANH, HOAI, HUONG) participated in the experiment. The pseudowords were presented one by one on a normal screen placed in front of the speakers. Five repetitions of each trisyllabic pseudoword plus the /aiaiaiaiai/ sequence, as well as five repetitions of the sentences, were presented in a random order and in two separated blocks.

Instructions given to the participants before the recording sessions began were to produce each item at a normal speaking rate and to pronounce a clear /i/ in the sequence /aiaiaiaiai/. This was verified before the recording sessions by asking them to pronounce the /aiaiaiaiai/ sequence once and then they were trained with 10 pseudowords (2 for each structure; with C=/f m k/, L=/R/, V=/a i/,) presented in a random order. Participants were shown the pseudowords on a computer screen one at a time using a powerpoint presentation with approximately 500ms intervals between slides (presentation was run manually by the experimenter).

The first results presented below are based on 50 CV.CVL. CVC tokens (2 pseudowords * 5 repetitions * 5 speakers) and 50 minus 2 /aiaiaiaiai/ tokens (5 repetitions * 5 speakers * 2 languages). Two repetitions of /aiaiaiaiai/ by two Vietnamese speakers were disregarded because pronounced [ajajajaja(j)] with a palatal glide instead of the high vowel expected.

2.3. Measures

Using semi-automatic segmentation with EasyAlign and Praat, the pseudowords were extracted and labeled with Trap v.6, a tool developed in GIPSA-lab by C. Savariaux in the Matlab® environment.

The following parameters were measured:

duration of the opening phase (corresponding to jaw lowering) and of the closing phase (corresponding to jaw raising), measured between successive maximum and minimum of the jaw trajectory (vertical component); these values were obtained by detecting the zero-crossing points on the velocity curve;

peak velocity and average velocity of each phase;

amplitude of each phase, which corresponds to the vertical displacement of the jaw, estimated between successive maxima of jaw opening and jaw closing.

2.4. Statistical Analysis

Our goal is to study variations in the response variables (duration, velocity and amplitude of each phase of the jaw cycle) as a function of syllable structure (CV, CVL, CVC, and the sequence of 10 vowels in the /aiaiaiaiai/ stimulus) and, for the /aiaiaiaiai/ stimulus only, as a function of language (French, Vietnamese).

Because our set of data contains repeated measurements made on the same speaker, we opted for linear mixed-effects models, and in order to respect the hypothesis according to which the residues follow a normal rule (condition for the application of mixed models), we chose to transform the response variable in its logarithm (ln, natural logarithm), thus:

$$\log Y_{ijkl} = \mu + \alpha_k + \beta_l + (\alpha\beta)_{kl} + \tau_i + \varepsilon_{ijkl}$$

in which:

Y_{ijkl}: is the j^{th} value of subject i taken by the response variable for modality k of factor A (syllable type), and l of factor B (language);
α_k: is the effect of modality k of factor A on the response variable;
β_l: is the effect of modality l of factor B on the response variable;
$(\alpha\beta)_{kl}$: is the effect of the interaction of modality k of factor A, and modality l of factor B, on the response variable;
τ_i: is the random effect of subject i on the response variable. follows a normal distribution in which is the response variable between the subjects.
ε_{ijkl}: represents the error term and follows a normal distribution.

To analyze differences between the two modalities (i.e. opening vs closing phases), the method of Hothorn et al. (2008) was applied in each modality of the syllable type and language, which allowed us to carry out multiple comparisons of means with the mixed model by ensuring that the risk of the first type related to all simultaneous decision making does not exceed the limit fixed in advance at 5% by adjusting the p-values. The method was applied to the data with the function glht of the package multcomp of the software R as well as the function lsmeans of the package lsmeans.

3. Results

3.1. Duration

The multiple comparisons of the average durations between the opening phase and the closing phase for each of the three jaw cycles found in the pseudowords CV.CVL.CVC (C={/t/, /s/}, V=/a/) and the five cycles of the sequence /aiaiaiaiai/ showed significant differences between the two phases, whereby closing is longer than opening (see Table 1, and Figures 1 and 2). An exception is found in the realization of the initial syllable /ta/.(t), for which the mean value of opening duration is 7ms greater than the value for closing. However, this difference is not significant. For this syllable, different patterns were observed among the five speakers (see Figure 3 for details), which explain the small and insignificant difference between the mean duration of the two phases.

Table 1. Estimations of mean differences (calculated after log-transformation from the linear model) between closing phase duration (C) and opening phase duration (O) for all the jaw cycles of /aiaiaiaiai/, /tataltat/ and /sasalsas/, with the standard deviation of the differences, the value of the statistics, and the p-value. In brackets are the onsets of the subsequent syllables.

	Estimate	Std. Error	z	p
C-O | iai	0.21906	0.04275	5.125	< 0.001 ***
C-O | sa.(s)	0.22734	0.06759	3.364	0.00537 **
C-O | sal	0.47024	0.06759	6.957	< 0.001 ***
C-O | sas	0.46185	0.06759	6.833	< 0.001 ***
C-O | ta.(t)	-0.06164	0.06610	-0.932	0.95156
C-O | tal	0.31084	0.06610	4.702	< 0.001 ***
C-O | tat	0.50349	0.06610	7.617	< 0.001 ***

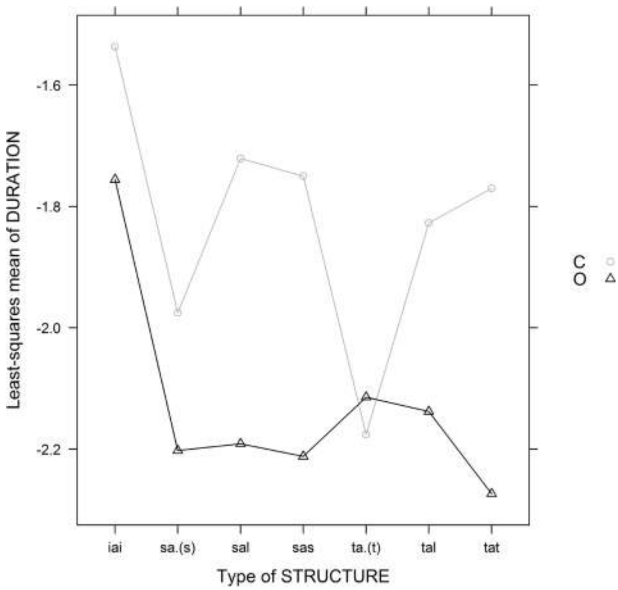

Figure 1. Estimation of average durations of opening and closing phases (log-transformed from the linear model) for each syllable type and /iai/ sequences. Closings are generally longer than openings. Least-Squares Mean is the mean estimated from the linear model.

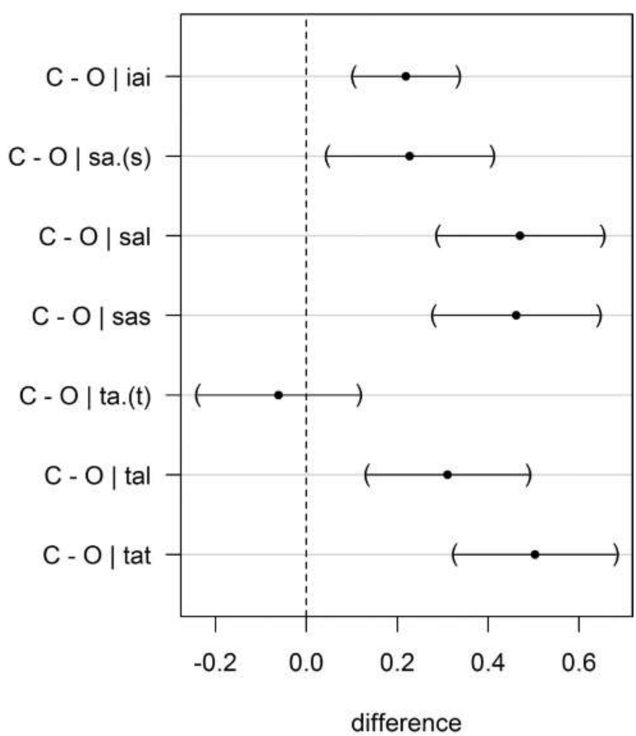

Figure 2. Point estimation of mean difference (from log-transformation) between closing phase (C) and opening phase durations (O) and 95% confidence intervals as a function of sequence type. Note that if the confidence interval includes zero, then the difference is not significant. Intervals are centered on the mean difference values, and their widths depend on variances and sizes. The wider the interval is, the more variability in the sample, the less accurate the estimate of the mean difference.

As can be seen in Figure 1, the raising of the jaw is of longer duration in the case of closed syllables. Against all odds, the longest phase durations (opening as well as closing) are observed in /iai/ sequences, perhaps because only vowels occur in the succession which requires more than one syllable gestures (in the sense of Articulatory Phonology). The comparison between jaw-closing phase durations of /iai/ sequences and that of open CV syllables indicates a shorter closing movement when the following segment is a syllable-initial consonant. The mean durations of jaw-opening phases exhibit less variation than closing phases according to syllable types. A lengthening effect in the pseudoword-final, most probably due to the French iambic accentual pattern, is only observed for the jaw-closing phase towards the alveolar plosive (/tat/).

Figure 2 shows the mean duration differences between the closing and the opening phases. Only for /ta/.(t) the difference is not significant. For the other sequences, the closing-to-opening ratio is positive; for the /iai/ sequence, the confidence interval is smaller, indicating less variability and therefore a more stable production compared to the other sequences.

Phase duration differences are in general smaller for the open syllables /ta/.(t) and /sa/.(s) than for the CVC types (Figure 3). Closing durations are longer for syllables with an offset consonant (on average 170ms vs 126ms). With the exception of /ta/.(t), the closing phase is thus consistently longer than the opening phase in the production of French speakers, regardless of syllable structure, type of the consonant, and position in the pseudowords (see Figure 3). Just a slight trend towards the reverse pattern (the opening phase being longer than the closing phase) is observed with ML speaker for /tal/. This unexpected result is not consistent with those of previous studies and especially as in CVC syllables initial C are most often longer than final C (Gracco, 1994; Redford, 1999).

The two phases of the jaw cycle were not related in their timing since no correlations were found between the opening and closing durations.

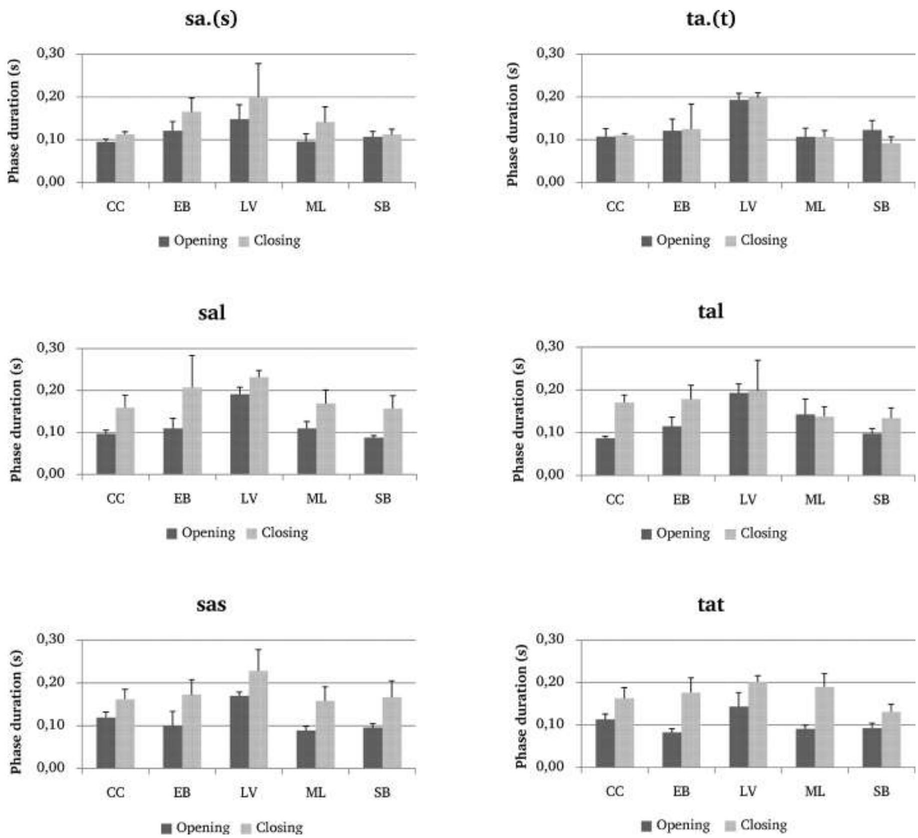

Figure 3. Mean durations of jaw-opening and jaw-closing phases according to syllable types, French female speakers, and both alveolar consonants /t/ and /s/.

3.2. Velocity

As in the case of duration (§3.1), velocity measurements of basic jaw movements reveal an asymmetric pattern. Measurements of the average velocity of opening and closing phases are in most cases significantly different (Table 2, Figure 4), with a faster opening and a slower closing (Figure 5). Only the initial open syllable /ta/.(t) does not follow this pattern showing similar average velocity for the two phases; the confidence interval in Figure 5 also indicates a large variability for this type of structure. Again, the smallest confidence interval is observed for /iai/ which is consistent with a more stable production pattern for this type of sequence compared to the others (Kelso

Table 2. Estimations of mean differences (calculated after log-transformation) between closing phase velocity (C) and opening phase velocity (O) for all the jaw cycles measured in /aiaiaiaiai/, /tataltat/ and /sasalsas/, with the standard deviation of the differences, the value of the statistics and the p-value.

	Estimate	Std. Error	z	p
C-O | iai	-0.14975	0.05811	-2.577	0.06772
C-O | sa.(s)	-0.26517	0.08620	-3.076	0.01458 *
C-O | sal	-0.59029	0.08620	-6.848	< 0.001***
C-O | sas	-0.44852	0.08620	-5.203	< 0.001***
C-O | ta.(t)	0.02152	0.29353	0.073	1.00000
C-O | tal	-0.38905	0.08431	-4.615	< 0.001***
C-O | tat	-0.33608	0.08431	-3.986	< 0.001***

et al. 1985). Differences between opening and closing are more important for the other syllables, and especially for CVL structures (/tal/, /tat/, /sa/.(s), /sal/ and /sas/).

Looking then at cross-individual variability (Figures 6 and 7), it can be shown that for the /ta/.(t) sequence, two speakers (ML and SB) have an inverse jaw velocity pattern (closing phase faster than opening) compared to the other speakers, while for the /tat/ sequence, only

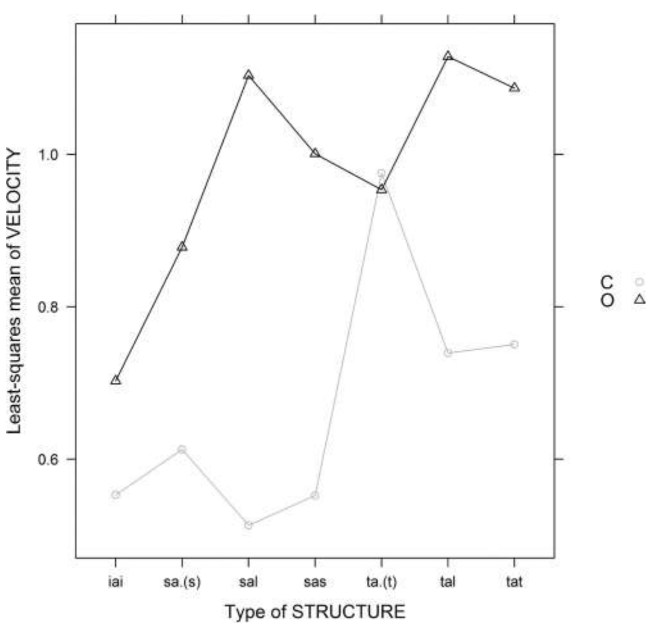

Figure 4. Estimations of the average velocity (log-transformed) of jaw-opening and jaw-closing phases as a function of sequence type. Except for /ta/.(t), the opening phases are faster than the closing phases.

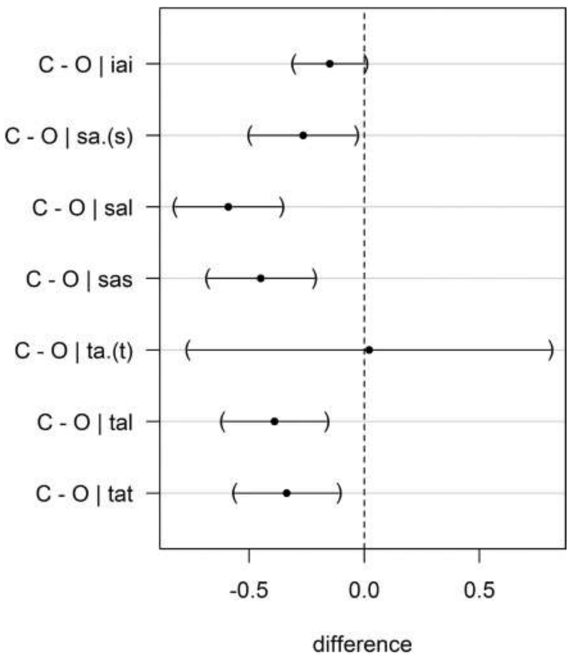

Figure 5. Point estimation of mean average velocity difference (from log-transformation) between jaw-closing (C) and jaw-opening phases (O), and the 95% confidence intervals as a function of sequence type. If the confidence interval include zero, then the difference is not significant.

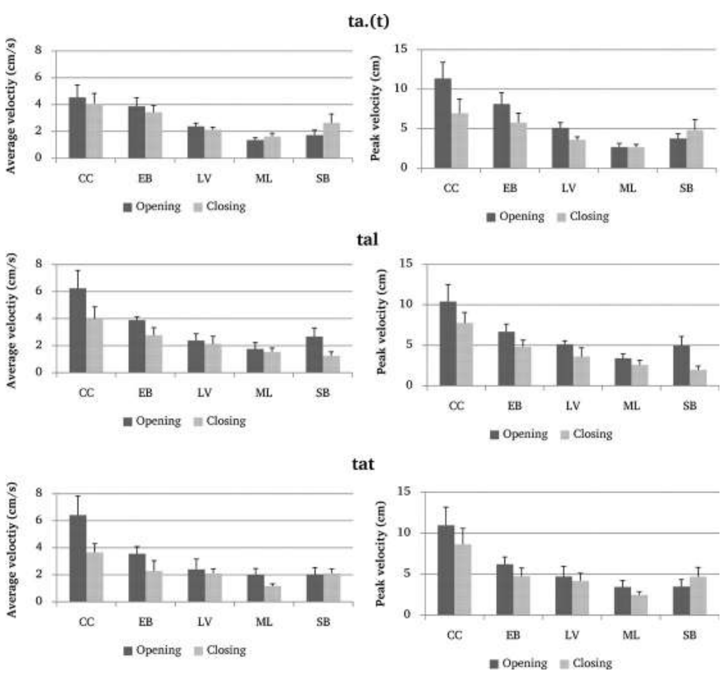

Figure 6. Mean values of average velocity (left) and peak velocity (right) of jaw-opening and jaw-closing phases for the syllables with /t/ across sequence types and speakers.

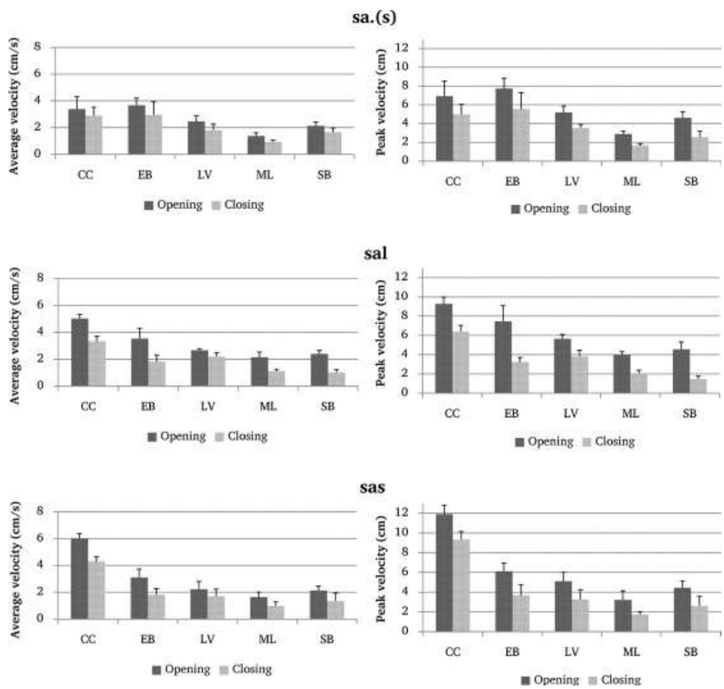

Figure 7. Mean values of average velocity (left) and peak velocity (right) of jaw-opening and jaw-closing phases for the syllables with the fricative consonant /s/ across sequence types and speakers.

one speaker (SB) has a slightly faster closing phase (Figure 6). The results for velocity peaks follow the patterns described for the average velocities. There were however no cross-individual differences for /s/-initial sequences (Figure 7).

For three of the speakers (CC, EB and LV), we found a statistically significant ($p < .05$) positive correlation between opening average velocity and closing average velocity (Figure 8), as well as between opening peak velocity and closing peak velocity for the sequence types (CV, CVL, CVC). This trend was not observed for the two speakers who showed a reversed velocity pattern for the initial syllable /ta/.(t): we found a weak positive correlation for ML and a negative correlation for SB, but statistically insignificant in the two cases ($p > .05$). When /ta/.(t) sequences were removed from the calculation, a statistically significant positive correlation was found between opening average velocity and closing average velocity for ML ($r = .45$, $p < .05$) but not for SB ($r = .27$, $p > .05$); the same observations were made for the velocity peaks.

In the case of the /aiaiaiaiai/ sequences for the five speakers, the r scores were positive, statistically significant and on average higher

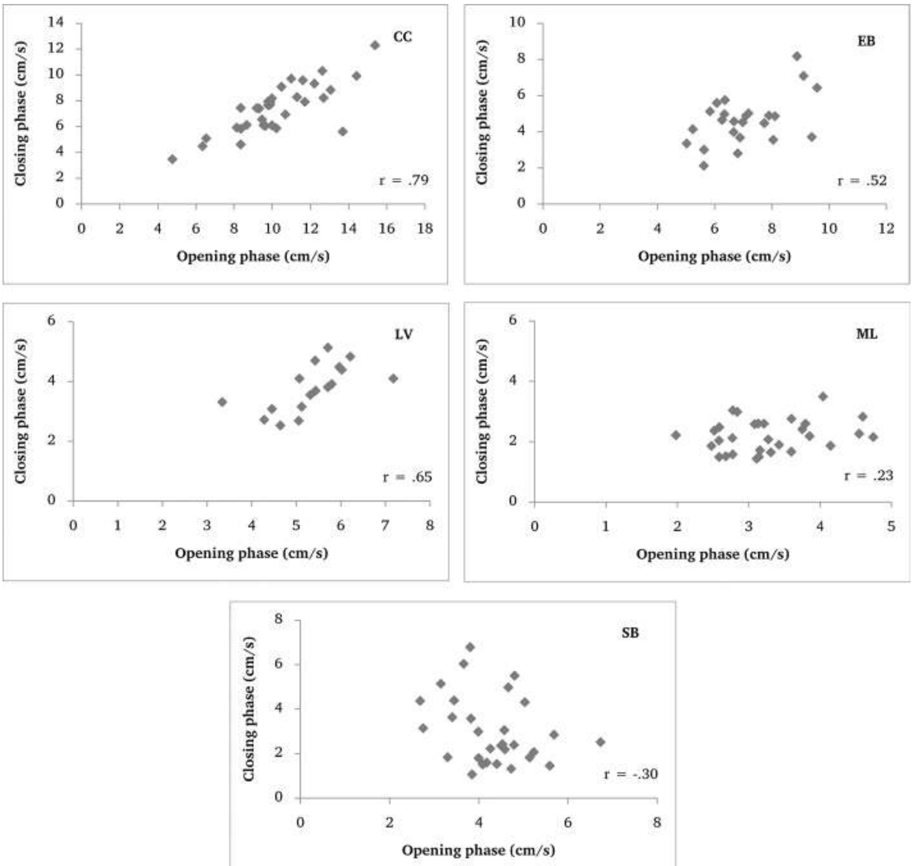

Figure 8. Correlations between jaw-opening and jaw-closing average velocities for each female speaker of French.

than for the other sequences ($r > .73$, $p < .05$). This result therefore suggests a more stable velocity pattern in the production of the vowel sequences.

For all the sequence types, the correlations suggest a relationship between closing phase velocity and opening phase velocity: the faster the opening the faster is the closing.

3.3. Amplitude

The amplitude, estimated as the vertical displacement of the jaw, shows no regular difference pattern between the opening and closing

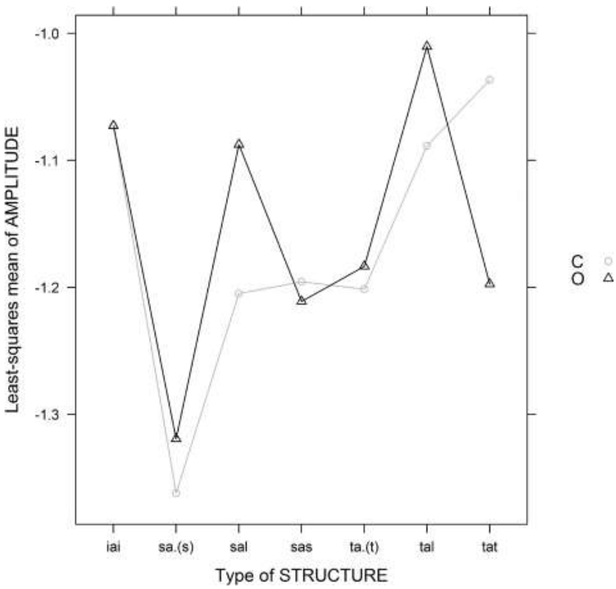

Figure 9. Estimation of average amplitudes of opening and closing phases (log-transformed from the linear model) across sequence types.

phases of the jaw cycles (Figure 9). Figure 10 shows that the differences in the average amplitudes between jaw-opening and jaw-closing phases are not significant for all the sequence types. When the speakers are individually analyzed, no consistency is observed between the differences of phase amplitudes. The same observation was made by Redford (1999) for CVC syllables. This result suggests that the amplitude values depend both on the segments and on the syllable structure types, but more data need to be analyzed to confirm this observation. Across speakers, however, there is a strong positive correlation between jaw-opening and jaw-closing amplitudes in the productions of /tataltat/ and /sasalsas/ sequences (CC: $r = .98$, EB: $r = .96$, LV: $r = .88$, ML: $r = .87$, SB: $r = .65$; for all $p < .05$), as well as in the production of the/aiaiaiaiai/ sequences (CC: $r = .97$, LV: $r = .83$, ML: $r = .88$, SB: $r = .73$; for all $p < .05$ and EB: $r = .70$, $p = .05$). The greater the displacement in opening phase the greater is the displacement in closing phase.

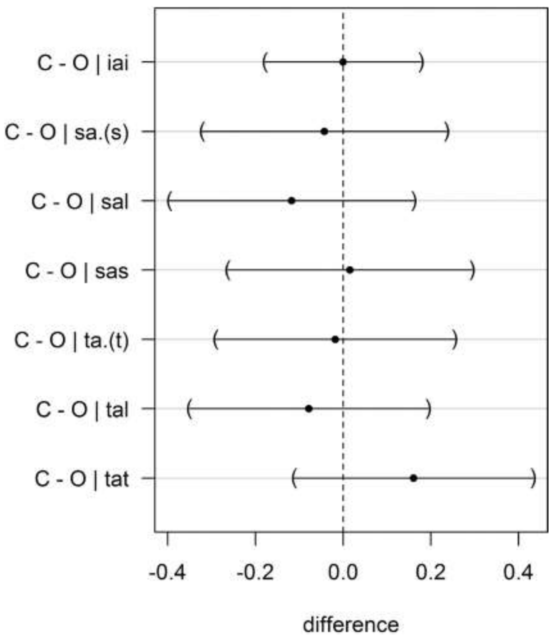

Figure 10. Point estimation of mean amplitude (from log-transformation) between closing phase (C) and opening phase durations (O), and the 95% confidence intervals for each type of jaw cycle. No differences are significant.

3.4. French-Vietnamese comparison

The cross-linguistic comparison only deals with the sequence /aiaiaiaiai/. The overall results of the Vietnamese speakers are similar to those of the French ones. The opening phases are shorter, and with lower amplitude than the closing phases (Figures 11). As for French, differences in durations and average velocities between jaw closing and jaw opening phases are significant (see Table 3 below).

Table 3. Estimations of mean differences of duration, average velocity, and amplitude (calculated after log-transformation) between closing phase (C) and opening phase (O) for all the jaw cycles measured in the sequence /aiaiaiaiai/ produced by the Vietnamese speakers, with the standard deviation of the differences, the value of the statistics and the p-value.

	Language	Estimate	Std. Error	z	p
Duration	C-O | Fr	0.21906	0.03722	5.886	< 0.001 ***
	C-O | Vn	0.16144	0.03067	5.264	< 0.001 ***
Velocity	C-O | Fr	-0.14322	0.05726	-2.501	0.0246 *
	C-O | Vn	-0.11455	0.04415	-2.594	0.0189 *
Amplitude	C-O | Fr	-0.00005	6.297e-02	-0.001	1.000
	C-O | Vn	0.0444	5.142e-02	0.864	0.625

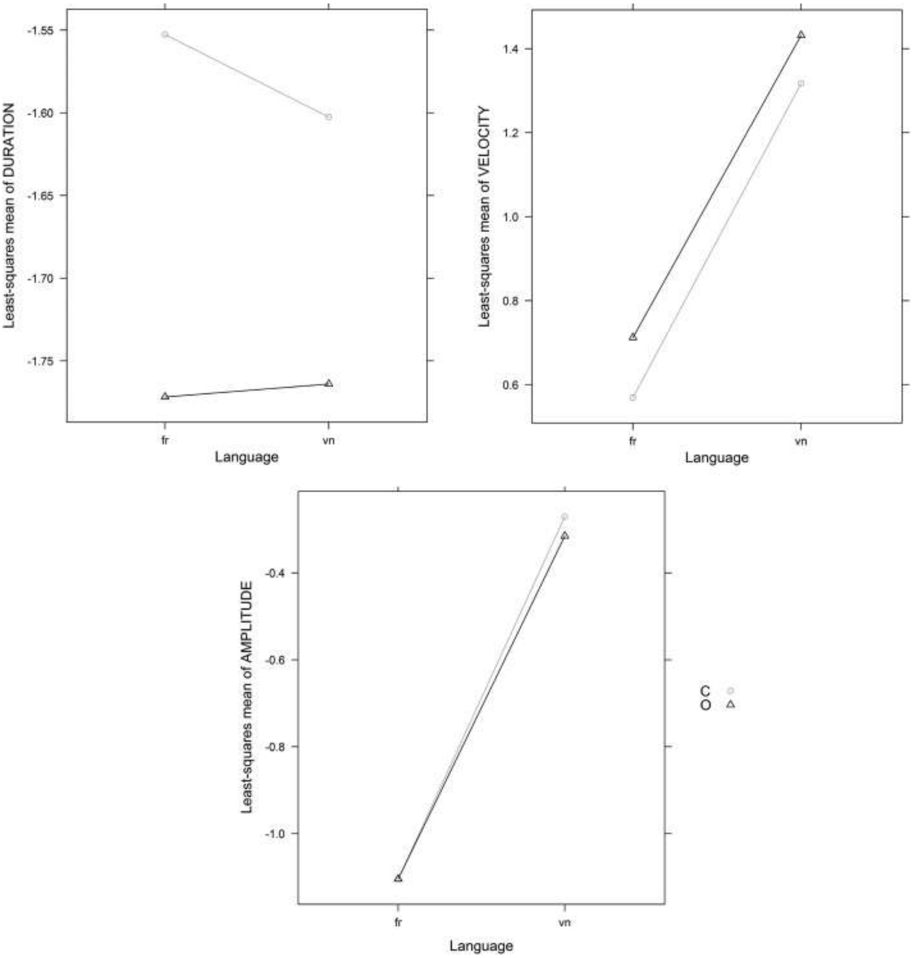

Figure 11. Estimation of (a) mean durations (b) mean velocities (c) mean amplitudes (expressed in log) of jaw-opening and jaw-closing phases for the sequence /aiaiaiaiai/ as a function of language.

4. Conclusion

The cross-language preference for syllable-initial over syllable-final consonants in speech, commonly known as the Maximum Onset Principle, is not clearly understood. The present study is part of an ongoing broader multilingual research undertaken to attempt to provide a phonetic basis for MOP. The relationship between syllable structures and basic open-close jaw movements were investigated within the framework of the Frame/Content Theory and following the findings of several previous studies in which asymmetric patterns in jaw oscillation cycles were observed. We have presented here the preliminary results obtained for French and Vietnamese.

For the two languages, our preliminary results confirm that there are asymmetries in the jaw cycle, with the opening phase (corresponding to the lowering of the jaw) generally shorter, faster, and with smaller amplitude than the closing phase (corresponding to the raising of the jaw). A few changes related to the context (stop vs fricative vs lateral, open vs closed syllables, pseudoword-initial vs final positions) were observed for duration and velocity, but they did not change the pattern of asymmetries. Measurements of the vowel sequences /aiaiaiaiai/ revealed a greater stability of the jaw cycles and stable patterns of asymmetry between the two phases for the duration, velocity, and amplitude parameters, most probably related to the reiteration of successive elements in the sequence (Kelso *et al.* 1985).

Although insignificant statistically, closing phases were found with smaller amplitude for the pseudoword-final syllables, which can be explained by the articulatory release observed in word endings. Several studies have shown that word-initial consonants have a stronger articulation than word-final consonants (for a review see Fougeron 1999; Krakow 1999). The latter are produced with less lingual contact and less linguopalatal pressure (Keating *et al.* 1999, Browman & Goldstein 1995). This word-edge effect in consonant articulation, and more specifically in the peaks of lingual gestures, was much larger for certain consonants and among them /t/, while for others, including /s/, only differences in acoustic duration were observed showing a greater duration for word-initial consonants. From these findings and since jaw supports the action of the tongue, we can expect an effect of word edge on jaw amplitude. However, we need more data to confirm this.

Following Gracco's suggestion that oral opening and oral closing are not independent actions, we calculated the correlations between the jaw opening phase and the subsequent closing phase for the duration, velocity, and amplitude parameters for each speaker. No correlations were found, in our study, between the opening and closing durations of the jaw by the fact that closing durations are different according to the syllable type and are linked to the presence of an offset consonant. Phase duration differences between opening and closing phases are in general smaller for open syllables. This result was expected since in CV types a single consonant (corresponding to the onset of the subsequent syllable) was articulated during the oral closing-opening movements while in CVC types two heterosyllabic consonants or a single pseudoword-final consonant (with a final lengthening) were articulated. Therefore, from our data, durations of the jaw closing phases seem linked to the syllable types (open vs closed). However, Redford & van Donkelaar (2008) found a lengthening of closing phases when a subsequent CC consonant sequence occurred as an onset (as in V.CCV) rather than as heterosyllabic (as in VC.CV), but this result was not easy to understand for the authors. The lack of correlation for durations of the two phases, observed in our data, is consistent with Gracco's observation that there were no influences of the opening durations on the closing durations of the jaw cycles. In Redford (1999), different patterns of duration in CVC syllable types were observed due to individual differences interpreted at a suprasegmental level (i.e. phrase final lengthening).

For all sequence types, positive correlations were found across speakers between jaw-opening and jaw closing velocities (averages or peaks) and between jaw-opening and jaw-closing amplitudes (displacements). The correlations, with coefficients ranging from r = 0.73 to r = 0.98, suggest that the two phases of the jaw cycles are strongly related following two tendencies: (1) the faster the opening the faster is the closing; (2) the greater the displacement in the opening phase the greater is the displacement in the closing phase. The fact that these tendencies were consistent across syllable types and across consonant types suggests that they are not necessarily related to the segmental phonological content. These results are consistent with those of Gracco (1994) who concluded that "[…] oral opening and oral closing are not independent events. Rather, oral closing actions are dependent on characteristics of the preceding oral opening action" (p. 16). He suggested that these findings came from biomechanical influences (i.e. muscle activity, tissue elasticity) on oral opening and oral closing movements of speech motor actions.

However, the asymmetrical patterns in the jaw cycle observed in our data with an opening phase generally shorter, faster, and with smaller amplitude than the closing phase contradict the previous works on American English conducted by Kelso *et al.* (1985), Gracco (1994), Redford (1999), Redford & van Donkelaar (2008). As mentioned in the introduction, these authors pointed out a reverse pattern, namely a longer, slower, and wider opening phase that was assumed to be the result of basic biomechanical properties of the jaw (lowering and raising are not subject to the same constraints). In our study, only the pseudoword-initial sequence /ta/.(t) shows an inverse pattern for both duration and velocity factors in the speech of the two French speakers ML and SB, but statistically insignificant.

This surprising result is not easy to understand. In the introduction, we suggested that consistent patterns of the jaw cycles showing longer and slower opening phases with greater displacements than the closing phases can support the hypothesis of a phonetic basis for MOP by the fact that longer opening and shorter closing durations can better accommodate the articulation of onsets (Gracco 1994). Such asymmetrical open-close movement of the jaw can account for the cross language preferences in syllable types for initial consonants or clusters over final consonants or clusters (also suggested by Redford 1999), as well as the preference for initial complex consonants over final complex consonants, and the larger proportion of Pure Frames in VC (and in vowel-coda of CVC) compared to CV structure. The reverse asymmetric pattern observed in our data is inconsistent with this reasoning. Thus a question to be solved in the future is why are jaw cycle patterns for French and Vietnamese speech exactly opposite to those observed for American English? The consistent pattern across French and Vietnamese excludes for the moment any explanations related to the basic syllable structure of the two languages: French is mostly a CV language while Vietnamese is mostly a CVC. Moreover and in contrast to French, Vietnamese disallows complexity in onset as in coda positions. However, it shows the asymmetrical pattern. Consequently and unlike our first hypothesis, the influence of the characteristics of the jaw cycle on syllable structure (as highlighted by the Frame/Content Theory) is not clarified here and requires further investigation.

However, in our study, the observed asymmetries in duration, velocity and, to a lesser extent (since no significant differences were found), amplitude of jaw cycles are in line with the assumption of Articulatory Phonology that vowels show different phasing patterns with onset and coda consonants. Indeed, a shorter period of time between the point of maximum jaw closing (consonant target) and the point of maximum jaw opening (vowel target) is consistent with a closer phasing of onset-consonant and vowel (in-phase relationship), and also with a high degree of coarticulation, or a synchronous temporal overlap, between C and V gestures. On the other side, a longer and slower jaw closing movement is compatible with a loose coupling between the nucleus and the following consonant(s). According to Articulatory Phonology, vowels and offset consonants are more time-independent (even in an anti-phase relation), and only the left-edge of a coda is phased with the vowel nucleus, making a coda more variable than an onset (Browman & Goldstein 1988, 1995; Byrd 1995). Within a short time interval it is obvious that intergestural phasing relations are stronger than within a longer time interval. In line with this reasoning, it is not clear how the reverse asymmetric pattern of jaw cycles observed in previous works on American English can be consistent with the basic AP predictions. Gracco (1994) suggested that onset consonants required oral opening and oral closing actions: the lengthening of the opening was helpful for VOT while the shortening of the closing phase supported a tight coordination between laryngeal and supralaryngeal movements. However, this explanation does not appear to account for our findings on French and Vietnamese. It is also inconsistent with Redford & van Donkelaar findings about V.CCV and VC.CV sequences: an unexpected lengthening of the jaw closing phase was observed in the V.CCV sequences. Further works will be required to clearly address this issue.

Results from the present study on jaw-opening and jaw-closing movements in French and Vietnamese speech which show opposite patterns to those observed in previous studies on American English encourage us to continue our multi-speaker and multi-language investigations, and to compare the current results with new data from native speakers of American English. The next step will be to investigate timing relations among lip, tongue and jaw movements, in relation with syllable boundary.

Addresses of the Authors

Nathalie Vallée, GIPSA-lab DPC, UMR 5216, CNRS & Université Grenoble Alpes, BP25 - 38040 Grenoble Cedex 9, France <nathalie.vallee@gipsa-lab.grenoble-inp.fr>

Thi Thuy Hien Tran, GIPSA-lab DPC, UMR 5216, CNRS & Université Grenoble Alpes, BP25 - 38040 Grenoble cedex 9, France <thi-thuy-hien.tran@gipsa-lab.grenoble-inp.fr>

Paolo Mairano, LFSAG, Università di Torino. Via Sant'Ottavio 20, 10124, Turin, Italy <paolomairano@gmail.com>

Silvain Gerber, GIPSA-lab, UMR 5216, CNRS & Université Grenoble Alpes, 11 rue des Mathématiques, BP46 38402 Saint Martin-D'Hères Cedex, France <silvain-gerber@gipsa-lab.grenoble-inp.fr>

Solange Rossato, LIG GETALP, UMR 5217, CNRS & Université Grenoble Alpes, Bât IMAG B, BP53 - 38041 Grenoble Cedex 9, France <solange.rossato@imag.fr>

Acknowledgements

This study is part of the project ANR-10-BLAN-1916 APPSy: Phonetic and Phonological Asymmetries of the Syllable. We would like to thank Christophe Savariaux and Quentin Tura for their valuable assistance in this project. Many thanks also to Myrna Laksman and Steve Huntley, and to the anonymous reviewers for their careful rereading and helpful comments on earlier versions of the paper.

Bibliographical References

Blevins Juliette 1995. The syllable in Phonological Theory. In Goldsmith John A. (ed.), *Handbook of Phonological Theory*. Oxford: Blackwell Publishers. 206-235.

Browman Catherine P. & Louis M. Goldstein 1988. Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45. 140-155.

Browman Catherine P. & Louis M. Goldstein 1990. Gestural Specification Using Dynamically-defined Articulatory Structures. *Journal of Phonetics* 18. 299-320.

Browman Catherine P. & Louis M. Goldstein 1995. Gestural Syllable Position Effects in American English. In Bell-Berti Fredericka & Lawrence J. Raphael (eds.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. New York: AIP Press: 19-33.

Browman Catherine P. & Louis M. Goldstein 2000. Competing Constrainsts on Intergestural Coordination and Self-organization of Phonological Structures. *Bulletin de la Communication Parlée* 5. 25-34.

Byrd Dani 1995. C-centers Revisited. *Phonetica* 52. 285-306.

Clements Nick & Samuel J. Keyser 1983. *CV Phonology: A Generative Theory of the Syllable*. Boston: MIT Press.

Davis Barbara L. & Peter F. MacNeilage 1995. The Articulatory Bases of Babbling. *Journal of Speech and Hearing Research* 38. 1199-1211.

Đoàn Thin Thut 1999. *Ngữ âm tiếng Việt* (Tr. *Vietnamese Phonetics*). Hanoi: Nhà xuất bản Đại học Quốc gia Hà Nội (Hanoi National University Publishing).

Fougeron Cécile 1999. Articulatory properties of initial segments in several prosodic constituents in French. *UCLA Working Papers in Phonetics* 97. 74-99.

Goldsmith John A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell LTD.

Goldstein Louis M., Byrd Dani & Elliot L. Saltzman 2006. The Role of

Vocal Tract Gestural Action Units in Understanding the Evolution of Phonology. In Arbib Michael (ed.), *From action to language: The mirror neuron system*. Cambridge: Cambridge University Press. 215–249

Gracco Vincent L. 1994. Some Organizational Characteristics of Speech Movement Control. *Journal of Speech and Hearing Research* 37. 4-27.

Hothorn Torsten, Bertz Frank & Peter Westfall 2008. Simultaneous Inference in General Parametric Models. *Biometrical Journal* 503. 346-363.

Kahn Daniel 1976. *Syllable-based Generalizations in English Phonology*. Bloomington: Indiana University Linguistics Club.

Keating Patricia, Wright Richard & Jie Zhang 1999. Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics* 97. 157-173.

Kelso J.A. Scott, Vatiokiotis-Bateson Eric, Saltzman Elliot L. & Bruce A. Kay 1985. A Qualitative Dynamic Analysis of Reiterant Speech Production: Phase Portraits, Kinematics and Dynamic Modeling. *Journal of the Acoustical Society of America* 77. 266-280.

Klein Marc 1993. La syllabe comme interface de la production et de la perception phoniques. In Laks Bernard & Marc Plénat (eds.), *De Natura Sonorum: essais de phonologie*. Saint-Denis: Presses Universitaires de Vincennes. 99-141.

Krakow Rena A. 1999. Physiological Organization of Syllables: A Review. *Journal of Phonetics* 27. 23-54.

MacNeilage Peter F. 1998. The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences* 21. 499-511.

MacNeilage Peter F.2008. *The Origin of Speech*. Oxford: Oxford University Press.

MacNeilage Peter F. & Barbara L. Davis 2001. Motor Mechanisms in Speech Ontogeny: Phylogenetic, Neurobiological and Linguistic Implications. *Current Opinion in Neurobiology* 11. 696-700.

Redford Melissa A. 1999. *An Articulatory Basis for the Syllable*. Ph.D. dissertation, The University of Texas, Austin.

Redford Melissa A. & Paul van Donkelaar 2008. Jaw Cycles and Linguistic Syllables in Adult English. In Davis Barbara L. & Krisztina Zajdo (eds.), *The Syllable in Speech Production: Perspectives on the Frame/Content Theory*. London: Taylor & Francis. 355-76.

Rousset Isabelle 2004. Structures syllabiques et lexicales des langues du monde: données, typologie, tendances générales et contraintes substantielles. Université Stendhal, Grenoble.

Selkirk Elisabeth 1982. The Syllable. In van der Hulst Harry & Norval Smith (eds.), *The Structure of Phonological Representations*. Dordrecht: Foris Publications. 337-383.

Tran Thi Thuy Hien 2011. *Processus d'acquisition des clusters et autres séquences de consonnes en langue seconde: de l'analyse acoustico-perceptive des séquences consonantiques du vietnamien à l'analyse de la perception et production des clusters du français par des apprenants vietnamiens du FLE*. Ph.D. dissertation, Université Stendhal, Grenoble.

Vallée Nathalie, Rossato Solange & Isabelle Rousset 2009. Favoured Syllabic

Patterns in the World's Languages and Sensorimotor Constraints. In Pellegrino François, Marsico Egidio, Chitoran Ioana & Christophe Coupé (eds.), *Approaches to Phonological Complexity*. Berlin: Mouton de Gruyter. 111-139.

Whalen Douglas H., Giulivi Sara, Goldstein Louis M., Nam Hosung & Andrea G. Levitt 2011. Response to MacNeilage and Davis and to Oller. *Language Learning and Development* 73. 243-2.