

The Labial-Coronal effect and CVCV stability during reiterant speech production: An articulatory analysis

Amélie Rochet-Capellan, Jean-Luc Schwartz

Institut de la Communication Parlée

INPG / Université Stendhal/ CNRS UMR 5009, Grenoble, France

amelie.rochet-capellan@icp.inpg.fr

jean-luc.schwartz@icp.inpg.fr

Abstract

In a companion paper [1], we showed that CVCV utterances with a labial consonant followed by a coronal one (LC sequences) are more stable than reverse CL sequences in speeded reiterant speech. We proposed that this could explain why human languages select LC sequences more often than CL ones (the “LC effect”). We provide here articulatory data explaining where the greater LC stability could come from, by investigating inter-articulator coordination during LC and CL utterances at an increasing rate. Rate increase leads variegated CVCV (e.g. /pata/) to be produced in a single jaw cycle but this is not the case for duplicated CVCV (e.g. /papa/. Furthermore, LC and CL sequences both evolve towards the same cycle with a progressive phasing of lips and tongue close together in the jaw cycle. Taken together, these results provide new elements to argue for motor control constraints shaping phonological patterns from economy principles.

1. Introduction

1.1. Regularities in language patterns

Comparative studies of languages phonology and human infants’ first productions have shown that *trends in the shaping of human languages mirror trends in the development of speech production* [1][2]. Firstly, CV (C for consonant and V for vowel) is the predominant syllable type in both canonical babbling and language inventories [3]. Secondly, CV combinations with the same place of articulation for the vowel and the consonant are preferred in both cases [2][4]. Thirdly, languages prefer to alternate place of articulation for successive consonants in CVCV disyllables, and in these variegated forms, Labial-to-Coronal patterns (/pata/, LC structure) occur more frequently than reverse order items (/tapa/, CL structure) [2][4]. Similarly, after reduplicated babbling (e.g. /papa/) progressively evolves towards variegation, infants display the same preference for “anterior to posterior progression” in their first words [2][5].

1.2. The “Frame then Content” (FC) theory

These observations lead researchers to suspect a common basis linked with basic properties of the perceptuo-motor system in relation to its ontogeny and phylogeny. The FC theory was elaborated in this framework [6]. It postulates a neo-Darwinian perspective on the birth of orofacial communication in both development and evolution, in which speech would have been elaborated from jaw movements preexisting in ingestion mechanisms. Thus, sequences of opening and closing jaw movements would constitute the

speech *frames*. Motor control of other articulators is later developed and evolves towards global coordination for the generation of *contents*.

1.3. Economy in substance-based phonological theories

The FC theory belongs to a general framework attempting to explain language universals from speech substance through general perceptuo-motor economy principles [7]. Focusing on the motor aspects, speech production requires energy; hence motor control programming should tend to select movement sequences limiting energy requirements. The first way to minimize the biomechanical cost of the articulatory system is to limit displacements of articulators. This could explain the CV co-occurrence patterns described above: sharing place of articulation for the consonant and the vowel would limit tongue displacements [2].

1.4. Economy, synergy and phase transitions in CVCV

Dynamic studies of motor systems have pointed out that rate increase may induce switches from one program to another. For example, minimizing energy consumption under velocity requirements may induce a horse to switch from alternating limb movements in trot to simultaneous movements in gallop [8]. Similar re-phasing between an organ and its counterpart in rate increase has been displayed for bimanual coordination [9].

Applying such principles of energy minimization, inter-articulator synergy, and phase transitions between various motor organizations could provide potential explanations for understanding the structure of CVCV sequences in the world's languages. Indeed, repeated motion of a single organ might reach its limit more quickly – and need a higher energy consumption – than achieving the same rhythm by the alternation of cycles of two organs coupled in phase opposition. This could mean that producing repeated motion with the same articulator (/papa/) would be less efficient and more costly than alternating articulator motion as in variegated CVCV (/pata/). This principle constraints piano playing: the same finger is never used twice for two consecutive notes.

Furthermore, for variegated sequences, inter-articulator synergy could depend on the degree of anticipation and phasing between articulators. In a companion study [1], we suggested that this could provide the basis for the LC effect. Indeed, we showed that during reiterant production of LC or CL CVCV at an increasing rate, LC sequences are stable, while CL sequences often switch towards the inverse LC pattern (e.g. /bada/ or /daba/ evolves towards a common attractor /badá/ or /bdá/). We interpreted this result as evidence for a greater stability of LC sequences compared with CL ones, possibly resulting from tongue-tip movement anticipation during the labial movement. This would allow a

better phasing of articulators for LC CVCV compare to CL and their grouping in a single jaw opening gesture.

In summary, motor economy would provide the basis for the LC effect. The present study tested this hypothesis. For this aim, we analyzed tongue-tip (TT), inferior-lip (IL) and jaw movement during reiterated utterances of LC and CL sequences produced by French speakers with a varying rate.

2. Method

2.1. Speech material

The phonetic material consisted of eight CVCV disyllable: two reduplicated ones: /papa/ and /tata/ and six variegated ones: three LC structures, /pata/, /pasa/ and /fata/ and their CL counterparts /tapa/, /sapa/ and /tafa/. The vowel /a/ was selected since it is a central unrounded vowel, requiring almost no active tongue or lip gesture. The selected labial and coronal consonants were varied in order to test the robustness of articulatory patterns observed over various manner of articulation.

2.2. Procedure

Five college-aged French speakers (three females and two males) without any speech or hearing deficits participated in the experiment. They were instructed to continuously repeat the disyllable enounced by the experimenter starting at a slow rate and then at an increasingly rapid rate up to whatever rate was possible. The speakers then had to progressively decrease rate to return to the initial slow rate. As verbal transformations (switch from one structure to another one [1][10]) could occur during this task, speakers were encouraged not to stop their productions even if they seemed different from the initial sequence. The eight disyllables were repeated three times with orders that varied from one subject to another.

Tongue-Tip (TT), Inferior-Lip (IL) and jaw displacements were tracked over time at 500 Hz using a Cartens Electromagnetic Articulograph. The acoustic signal was simultaneously recorded by the way of a microphone fixed on the articulograph helmet and then digitized at 20 kHz. This resulted in 18 wave files per subject (three per sequence) and the 18 associated articulatory binary files.

2.3. Data analysis: assumptions and measurements

The first hypothesis was that for slow rates, each CVCV production would respect a general rule: one jaw cycle for one syllable. Then, biomechanical properties would limit the decrease of jaw cycle duration above a floor value [11]. Disyllable duration would not decrease under twice this value for duplicated CVCV. However, shorter disyllable durations should be reached for variegated CVCV, due to the possibility of phasing both the labial and the coronal gesture in a single jaw cycle. Finally, progressive rephasing would happen between the labial and the coronal gesture (as observed before, [1]); hence the interval from IL constriction to TT constriction might be shorter than the one from TT constriction to the next IL one.

To test these hypotheses, the intensity curve computed from the acoustic signal was first labeled in the same way as in the companion study [1] in order to compare the intensity of V_L (the vowel following the labial consonant) and V_C (the vowel following the coronal consonant). This measure allows

us to track vowel overshoot and reduction and to infer whether the speaker is evolving towards LC (/patá/→/ptá/) or CL (/tapá/→/tpá/); V_C is more intense than V_L in the first case, and less so in the second one. Vowels were labeled as energy maxima and consonants as minima. Syllable and disyllable durations were measured from the acoustic signal as respectively, the time between two successive consonants and the time from the first consonant to the first consonant of the next disyllable.

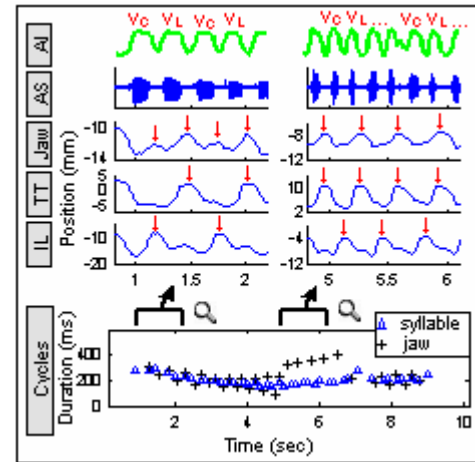


Figure 1: Example of /tapa/ utterances. Bottom: durations of jaw and syllable cycles across time. Top: sample of articulator trajectories, acoustic signal (AS) and intensity curve (AI) for slow rate (left) and fast rate (right). Arrows point to detected maxima (see text for details).

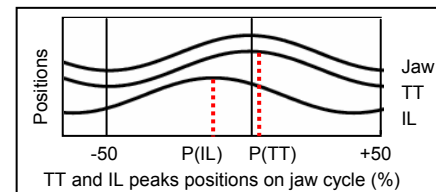


Figure 2: Phasing measurement between TT, IL and jaw. A closing then opening jaw gesture constitutes a jaw cycle normalized between -50% and 50%, 0 being the highest position. $P(IL)$ and $P(TT)$ are respectively the position of the labial and coronal constriction relative to jaw cycle.

The acoustic labeling was then used to guide articulatory processing. Examples of articulator trajectories are displayed in fig. 1 (top). Articulatory trajectories were first preprocessed by a low-pass Chebychev filter. TT constrictions were then labeled as the highest TT positions between V_L and V_C , and IL constrictions as the highest IL positions between V_C and V_L . Thus, secondary maxima due to articulator passive motion due to jaw gesture for the constriction of the other articulator are not labeled (Fig. 1, top left). Jaw maxima and minima were also detected from the jaw trajectory. Final post-processing discarded spurious events and too small jaw movements (strokes with amplitude lower than 0.1 mm). Values were stored in binary files. This allowed the following values to be compute:

- The duration of the mandible cycle,
- An indicator of phasing between TT and IL: the duration of the labial-to-coronal phase (time from IL constriction to TT constriction D_{LC}), normalized by the whole disyllable duration (time from IL constriction to the next one D_{LL}):

$$\Delta_{(LC/LL)} = D_{LC} / D_{LL}$$

- The position of each constriction event relative to the jaw cycle expressed in percents of cycle (Fig. 2).

3. Results

Acoustic data were analyzed in the same way as in the companion study. The results are quite coherent for the five speakers of the present study, and the 22 speakers of the companion study. They display similar variability between speakers and sequences, together with greater stability of LC sequences as compared to CL ones. Due to space constraints, these results could not be presented here, and articulatory analysis will focus on /papa/, /tata/, /pata/ and /tapa/ CVCV, globally representative for other sequences.

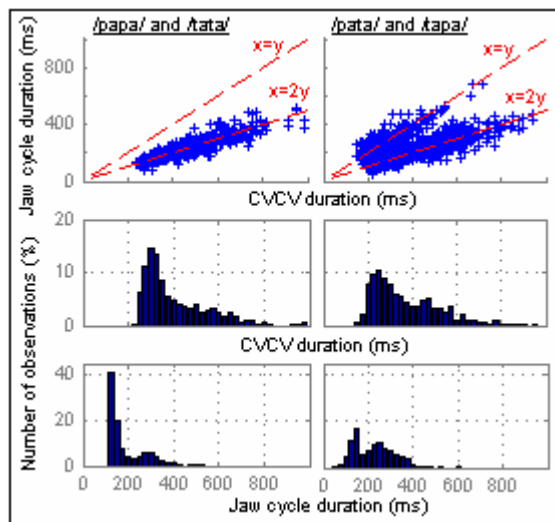


Figure 3: Relationship between disyllable and jaw cycle duration, grouping all speakers' utterances. Left: duplicated CVCV. Right: variegated ones. Top: jaw cycle duration against disyllable duration. Middle and bottom, respectively: histograms of disyllable and jaw cycle durations.

3.1. Disyllable vs. jaw cycle duration

Jaw cycle duration and syllable duration plotted against time show that during the acceleration phase, durations progressively decrease together (Fig.1, bottom row). Then, when the syllable rate reaches around 200 ms, the jaw cycle duration separates from the syllable duration, and rather suddenly doubles in length. This indicates that the two syllables are now realized in a single jaw cycle. The global analysis displayed on Fig.3 confirms this observation. Indeed, the plot of jaw cycle duration against disyllable duration shows that productions for duplicated CVCV are grouped around an $(x=2y)$ line, with two jaw cycles for one disyllable. However, variegated productions are

distributed both around the $(x=2y)$ and $(x=y)$ lines. Values around the $(x=y)$ line indicate that there are a number of sequences with two syllables for only one jaw cycle (Fig. 3, top row). For all sequences, jaw cycle is almost never shorter than 100 ms. Therefore, durations of duplicated disyllables are never shorter than 200 ms and peaks around 300 ms. For variegated sequences, however, it can reach values under 200 ms, and peaks around 200-250 ms (Fig.3, middle row). Finally, the comparison of jaw cycle histograms for duplicated and variegated CVCV (Fig. 3, bottom row) shows that there are two peaks in both case, one around 125 ms, the other around 250 ms. However, the peak at 250 ms is larger for variegated than for duplicated sequences.

3.2. Phasing between jaw and constrictors

The switch from a one-to-one towards a one-to-two relation requires a reorganization of inter-articulator coordination. Indeed, when each syllable is realized in its own jaw cycle, TT and IL are both phased with the jaw, occurring at a high jaw position (Fig.1, top left). Then, with rate increase, the two constrictors may enter into a single jaw cycle. The plots in Fig.4 display how this reorganization occurs. It appears that for four out of five speakers (S1, S3, S4, and S5), the labial constriction position peaks at -37.5% for both /pata/ and /tapa/, (Fig.4, top), while the coronal constriction peaks around 0% . The result is different for speaker S2 who realizes /p/ around -12.5% and /t/ around 25% . Altogether, the labial occurs in the raising phase of the jaw cycle, and the coronal around the jaw peak or slightly after, with a temporal distance between the labial and the coronal constriction lower than half a jaw cycle. The synchrony of the coronal constriction with the high position of the jaw within its cycle is also observed for /pasa/, /sapa/, /fata/ and /tafa/ (not presented here).

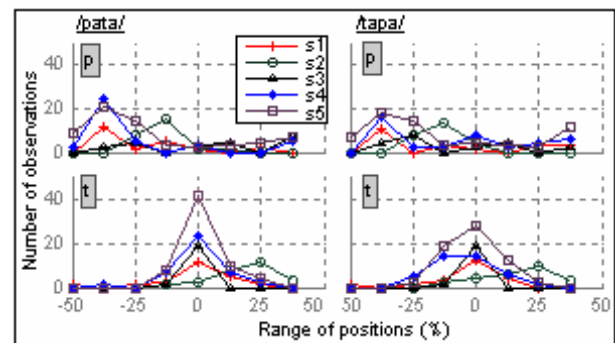


Figure 4: Position of IL constriction (/p/, top) and TT constriction (/t/, bottom) on the mandible cycle for /pata/ (left) and /tapa/ (right) when two syllables are grouped inside a single jaw cycle.

3.3. Phasing between TT and IL

Table 1 presents statistics on the duration of the labial-to-coronal phase relative to the disyllable duration, $\Delta_{(LC/LL)}$. Values of $\Delta_{(LC/LL)}$ lower than 50% indicate that /p/ is closer to /t/ than /t/ is to the next /p/, hence a greater synchrony for /pt/ than for /tp/. Four out of five values are significantly lower than 50% for /pata/, and three out of five for /tapa/.

Speakers	/pata/			/tapa/		
	Mean	std	Nb	Mean	std	Nb
S1	49.35	6.16	20	53.05	2.64	5
S2	36.32*	9.30	25	40.34*	11.26	21
S3	44.3*	5.70	19	38.61*	5.7	11
S4	45.41*	9.10	56	49.46	12.56	48
S5	42.08*	8.05	67	40.84*	8.64	58

Table 1: Means, standard deviations and number of observations of $\Delta_{(LC/LL)}$ for /pata/ and /tapa/ per speaker when two syllables are grouped inside a single jaw cycle. Asterisks indicate means significantly different from 50% at the 0.01 level after Bonferroni correction ($p \leq 0.01/5$).

4. Discussion

Taken together, the results are in line with our expectations. Furthermore, they agree with previous studies on jaw motion and provide new elements to argue for motor control constraints shaping phonological patterns due to economy principles.

4.1. Jaw saturation

Jaw cycle durations display two peaks around 125 ms and 250-300 ms (Fig.3, bottom). Similar durations were found by Nelson et al. [11] for "sa-sa" utterances. These authors pointed out that 50 ms is the minimum duration of jaw strokes (closing vs. opening movement, corresponding to half of our jaw cycle). This is compatible with the first peak at 125 ms and the duration floor at 100 ms in our data. Moreover, their results displayed a "resonance" phenomenon: maximum motion amplitude was observed for stroke durations around 125 ms with an acceleration peak lower than for smaller duration cycles. This indicates smaller energy consumption, and it could explain the second peak of jaw cycles around 250-300 ms in our results. To summarize, 100 ms cycle duration corresponds to the saturation of the jaw oscillatory system. Speakers can reach such rates for both variegated and reduplicated sequences but it is more costly in energy than the slower preferred rate around four cycles per second. This would lead speakers to switch their motor program for variegated utterances from one cycle per syllable to one cycle per disyllable. This shift induces increasing proximity of the lip and tongue constriction to allow fast utterances.

4.2. Alternated vs. reduplicated motions

Indeed, realization of two syllables on a single jaw cycle could be possible for variegated LC and CL CVCV because coronal constriction could be anticipated during labial constriction while the inverse tendency would be rarely observed. That suggests a progressive evolution from both LC and CL structures to the LC pattern with an undershoot of V_L . However, while a TT movement could be initiated during an IL one, such anticipation is impossible for reduplicated CVCV: a single articulator could not begin a new movement if it has not finish the preceding one. Thus, alternating movement of two different articulators allows the attainment of fast rate production with little energy consumption. Coupled with asymmetries in phasing between IL and TT, that provides an explanation of both the LC effect and the dominance of variegated forms in lexicons of world's languages in term of economy of energy.

5. Conclusions

The preference for variegated over reduplicated patterns and for LC over CL sequences seems to be possibly related to motor economy constraints. Nevertheless, the greater articulatory stability of LC sequences also has some perceptual counterparts, as shown by the greater stability of LC over CL sequences in verbal transformation experiments [12]. More generally, motor economy should be combined with perceptual stability and distinctiveness criteria [13] to progress towards a general substance-based theory of the patterning of sound systems in human languages.

6. Acknowledgements

This work is part of the "Patipapa" project funded by the French Ministry of Research (Action Concertée Incitative "Systèmes Complexes en Sciences Humaines et Sociales").

We also thank C. Savariaux and C. Vilain for their technical support and the five courageous speakers for their essential participation.

7. References

- [1] Rochet-Capellan, A. and Schwartz, J.-L. "The Labial-Coronal effect and CVCV stability during reiterant speech production: An acoustic analysis", this conference submitted.
- [2] MacNeilage, P.F. and Davis, B.L. "On the origins of internal structure of word forms", *Science*, 288:527-531, 2000.
- [3] Davis, B.L., MacNeilage, P.F. and Matyear, C. "Acquisition of serial complexity in speech", *Phonetica*, 59(2-3):75-107, Apr-Sep 2002.
- [4] Rousset, I. "From lexical to syllabic organization: Favored and disfavored co-occurrences", *Proc. XVth ICPHS, Barcelona*, pp. 715-718, 2003.
- [5] Locke, J. "Movement patterns in spoken language", *Science*, 288:449-451, 2000.
- [6] MacNeilage, P.F. "The frame/content theory of evolution of speech production", *Behavioral and Brain Sciences*, 21:499-511, 1998.
- [7] Lindblom, B. "On the notion of possible speech sound", *Journal of Phonetics*, 18, 135-152, 1990.
- [8] Hoyt, D. and Taylor, C.R. "Gait and the Energetic of Locomotion in Horses", *Nature* 292, 239-240, 1981.
- [9] Haken, H., Kelso, J.A.S., and Bunz, H. "A theoretical model of phase transitions in human hand movements", *Biological Cybernetics*. 51:347-356, 1985.
- [10] Warren, M.R. "Illusory changes of distinct speech upon repetition – The verbal transformation effect", *British Journal of Psychology*, 52: 249-258, 1961.
- [11] Nelson, W.L., Perkell, J.L. and Westbury, J.R. "Mandible movements during increasingly rapid articulations of single syllables", *J. Acoust Soc Am*, 75(3):945-951, 1984.
- [12] Rousset, I., Sato, M., Schwartz, J.-L. and Vallée, N. "Un corrélat perceptif de l'effet LC", *Actes des XXVèmes Journées d'Etudes sur la Parole*, 441-444, 2004.
- [13] Schwartz, J.L., Boë, L.J., Vallée, N., and Abry, C. "The dispersion-focalization theory of vowel systems", *J. of Phonetics*, 25, 255-286, 1997.