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

Amélie Rochet-Capellan, Jean-Luc Schwartz

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

This study investigates the stability of LC (Labial-Coronal) and CL (Coronal-Labial) CVCV sequences during a repetition task at an increasing rate. Despite variability between subjects and differences linked to the various consonant manners of articulation used in the study, the results show that the LC pattern is more stable than the CL pattern. Indeed, under the rate pressure, CL sequences often evolved to LC ones (e.g. /sapa/ switched to /pasa/) whereas the reverse pattern almost never happened. Coupled with previous studies about world languages, speech development and verbal transformation experiments, this is interpreted as a possible influence of the substance on phonological forms.

1. Introduction

1.1. The LC effect in phonology and development

Statistical analysis of world languages have shown that phonological systems use only a little set of phonemes and phonemes combinations among a great number of possible selections. Firstly, syllables are mainly of the CV type (C for consonant and V for vowel: e.g. /pa/ is preferred to /ap/ [1]). Moreover, co-occurrences of C and V sharing the same place of articulation are the prevailing intrasyllabic patterns (e.g. /di/ or /gu/ are preferred to /du/ or /gi/ [2]). Then, at an inter-syllabic level, CVCV that variegate the articulator from the first to the second syllable are more often used than reduplicated sequences (/pata/ rather than /papa/). Among these possible variegated forms, sequences that begin with a labial (L) consonant followed in the second syllable by a coronal (C) consonant are more likely than the reverse pattern (e.g. /pata/, an LC structure, rather than /tapa/, a CL one). This asymmetry, called the Labial-Coronal (LC) effect, seems strong (with 2.5 times more LC than CL sequences in world languages [2][1]).

These trends in the *shaping of human languages* mirror trends in the *development of speech production*. Indeed, the earliest phase in canonical babbling is characterized by repeated CV sequences (reduplicative babbling) with C and V more often sharing the same place of articulation. Reduplicated babbling is then progressively transformed into variegated one to finally reach the first word developmental step. Emergence of the LC effect is a strong characteristic of this last transition [3].

1.2. Hypotheses about the LC effect

Such similarities have lead researchers to suspect that the selection of speech forms in babbling and world languages could result from the same substance-based constraints. This common basis might be linked with basic properties of the perceptuo-motor system and its development in the course of ontogeny and phylogeny. This is the rationale of the "Frame then Content" (FC) theory [2], according to which jaw is the first articulator to be active in canonical babbling, producing *frames*, that is alternations of close and open patterns called closants and vocants. These madibular oscillations carried the other passive articulators resulting in articulation co-occurrences between closants and vocants. Then, *content* could progressively emerge from the development of independent motor control of all the articulators. The LC effect in this context is supposed to be due to a "simple first" mechanism linked with the assumed tendency of motor systems to initiate voluntary actions by the simplest movement. The labial closants would be simpler than coronal ones because they are supposed to be produced with no active tongue movement (hence called "pure frames") whereas coronal closants would induce an active tongue movement (hence called "fronted frames"). The latter are supposed to be easier to produce secondly in the utterance [3].

This explanation of the LC effect raises some problems for two reasons. Firstly, the "simple first" explanation seems a bit *ad hoc*. Secondly, both simulations on vocal tract models and canonical babbling data indicate that pure frames could in fact involve coronals as well as labials [4].

This lead us to propose another assumption, according to which LC sequences are preferred because they could be produced inside a single jaw opening gesture, the tongue-tip movement being anticipated during the labial movement, while the converse would not be possible. This would lead speakers to organize CVCV of the LC type into a single global gesture, while it could not be the case for CL sequences. A first confirmation of this idea was obtained in a verbal transformation experiment [5]. In this paradigm, the sound stream created by rapid and continuous repetition of a word could give rise to different speech percepts [6]. A study of verbal transformations of /pVtV/ or /tVpV/ sequences showed that listeners perceived both types of sequences more often as LC forms than as CL ones. This was interpreted as a perceptual correlate of the LC effect. It would result from the greater perceptuo-motor stability of the LC pattern due to the degree of articulatory phasing between the labial and the coronal gestures.

1.3. An experimental test

The **s**tability of phonologic forms can also be investigated by the reiterant speech paradigm: repeated speech with increased rate would reveal the most stable modes of coordination of the articulatory system [7]. Using this paradigm, experimental studies have shown that fast repetition rates could induce resyllabification from a VC form (like /ib/) to a CV form (/bi/) (from Stetson, 1951 [8]; to de Jong, 2001 [9]). CV syllables, more stable than VC forms in articulatory terms, would therefore be preferred in human languages. Our investigation used this paradigm to test for a possible LC preference in speech production at increasing rates. If LC sequences are better phased and hence more stable than CL ones, increasing the production rate should lead the speaker to select the most stable sequence. Consequently, LC utterances should remain stable whereas CL ones should transform into the reverse LC schema. For example, /pata/ and /tapa/ would both change for /patá/ and then for /ptá/ with the complete phasing of gestures of the lips and the tongue tip at the beginning of the opening gesture.

2. Method

2.1. Phonetic material

The phonetic materiel consisted of six CVCV: three LC structures, /pasa/, /bada/ and /vada/ and their CL counterparts /sapa/, /daba/ and /dava/. 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 a possible LC preference over various consonant modes.

2.2. Procedure

Speakers were 29 undergraduate students of the University of Grenoble, all native French speakers without any hearing problems and naive to the purpose of the experiment. The instruction was to repeat sequences proposed by the experimenter, starting in a slow way and progressively increasing the rate to reach the level described by the instruction: "as fast as you can". Speakers were encouraged not to stop their productions even if it seemed different from the initial sequence (that is if any verbal transformation occurred). Sequences were proposed in counterbalanced orders from one subject to the other. Productions were tape-recorded and then digitized at 22 kHz. This resulted in a set of twenty-nine wave files, each made up of six utterances, one per bi-syllable.

2.3. Data analysis

The hypothesis was that LC sequences would be more stable than CL ones under rate pressure. This stability would be due to a progressive synchronization of the labial and coronal gestures possibly resulting into a single consonantal cluster. This would induce an increasing undershoot of the vowel following the labial consonant (V_L in the following) until its possible complete fading out. Let 'LaCa' be the general representation of LC sequences. The speeding process could be: LaCa→LaCá→LCá (/pasa/→/pasá/→/psá/). Similarly, utterances for CL sequences could be: CaLa→LaCa→LaCá→LCá (/sapa/→/pasa/→/pasá/→/psá/). This process could be displayed through prosodic measurements implying either the fundamental frequency, or the syllabic duration, or the vowel intensity. We focused here on vowel intensity, which provides a cue both sensitive and easy to process semi-automatically.

In order to select a continuous utterance for each sequence, all files were edited with the Praat software [10]. One wave file was created per sequence for each subject. Then, using a Praat script, consonants and vowels were hand-labeled on intensity plots as, respectively, local minima and maxima of the energy curve (Fig. 1). When speakers switched to a single syllable (e.g. /psa/) labeling rule was to mark the "vanished" vowel and the two consonants around it as a same event (Fig.1, right). Labeling was automatically stored in a text file.

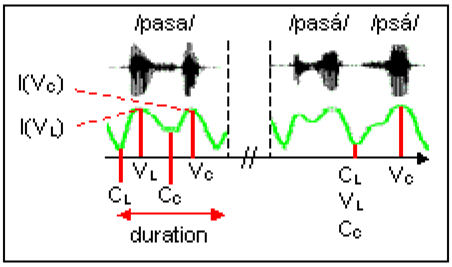

Figure 1: Data processing: sample of /pasa/ utterances. Top: signal, Bottom: intensity against time. C_L and C_C are the labial and coronal consonants; V_L and V_C the vowels respectively following C_L and C_C; $I(V_L)$ and $I(V_C)$ their intensity.

From these data, two parameters were computed for each utterance of each sequence (displayed in Fig. 1):

- The utterance duration: temporal distance between the first consonant of the current utterance and the first consonant of the next utterance,
- The difference between the intensity of the vowel after the coronal consonant, V_C, and the intensity of the vowel after the labial consonant, V_L:

$$\Delta I = I(V_C) - I(V_L)$$

The prediction was that ΔI would be positive for both LC and CL sequences for fast productions.

3. Results

3.1. Global tendencies

The values of ΔI as a function of utterance duration are plotted on Fig.2 for all utterances and all speakers, for each of the six sequences. The values are grouped around 0 for durations above 250-300 ms. So, for slow rates, the speakers

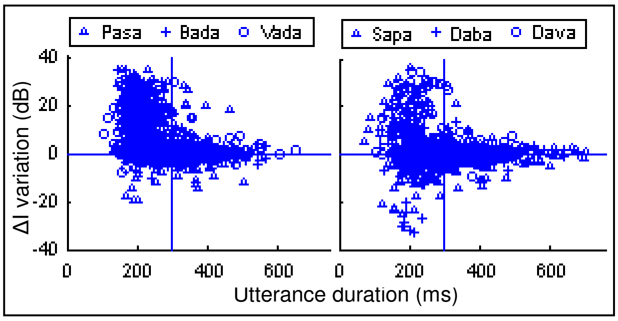

Figure 2: Intensity variation between V_C and V_L (ΔI) against utterance duration for all subjects. Left: LC sequences, right: CL sequences

keep a regular alternation of labial and coronal syllables with no strong accentuation or reduction effect. Below this boundary around 250-300 ms, the range of ΔI values increases for both LC and CL sequences. In the LC group (left part of the figure), most of the extension is towards positive ΔI values, which means that the evolution is done towards LaCá and then LCá. In the CL group the situation is more complex. Though ΔI is greater than zero for most productions, a number of negative values occur. This indicates that though most shifts occur towards LaCá and LCá, some occur in the inverse direction, towards CaLá and then CLá.

3.2. Detailed analyses

In order to compare LC and CL stability under rate pressure, and considering that according to Fig.2., the most interesting patterns happen for utterance durations below 250-300 ms, analysis were limited to this low-duration range. However, some speakers never reached such small durations. To obtain analyses as homogeneous as possible, we kept only the data from speakers who produced at least five utterances with durations below 300 ms for each of the six sequences. This criterion excluded seven subjects that were too slow for the experimental paradigm.

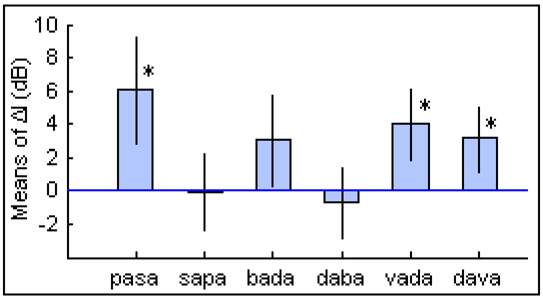

Figure 3: ΔI means and standard errors for each sequence, averaged over all subjects and all utterances with duration below 300 ms. The stars indicate that the t-test reached significance at the 0.05 level after Bonferroni correction (p≤0.05/6).

In the following, only the productions with durations less than or equal to 300 ms are analyzed, for the 22 remaining subjects. Means and standard deviations for ΔI are displayed for each sequence on Fig.3. Six repeated two-tails Student tests with Bonferroni correction were performed to test if each mean significantly differed from zero. The same analyses were done for each individual subject, resulting in 22*6 two-tails Student tests with Bonferroni correction. Histograms of ΔI means together with results of the statistical analyses for each sequence are displayed in Fig.4. The results will now be discussed separately for each type of LC vs. CL contrast.

Pasa/sapa

The ΔI mean averaged over all utterances with duration under 300 ms and over all speakers is positive for /pasa/ (6.0dB) and almost 0 for /sapa/ (Fig.3, left). The two-tail Student test is significant for /pasa/ (p=0.002). The analysis per subject confirms the trend for /pasa/, with significant positive ΔI means for five subjects, and no significant negative value. For /sapa/, though the histogram of mean ΔI values is quite symmetric around 0, means are significantly different from 0 in four cases for negative values, but never for positive values (Fig. 4, first row). Altogether, the trend is to go from /pasa/ to /pasá/ and /psá/, and from /sapa/ either to /sapá/ and /spá/ or to /pasá/ and /psá/ depending on subjects.

Bada/daba

The pattern for /bada/ and /daba/ is similar to the previous one. The global ΔI mean (Fig. 3, middle) is positive for the LC sequence /bada/ (3.1dB), though not significantly different from 0; and close to 0 though weakly negative for the CL sequence /daba/ (–0.6dB). The repartition of individual means (Fig.4, second row)) confirms this portrait. For /bada/, five subjects display significantly positive ΔI mean. For /daba/, only one negative ΔI mean is significantly different from zero. Altogether, the LC structure /badá/ or /bdá/ provides the attractor for /bada/, while /daba/ utterances change for /dabá/ as well as /badá/.

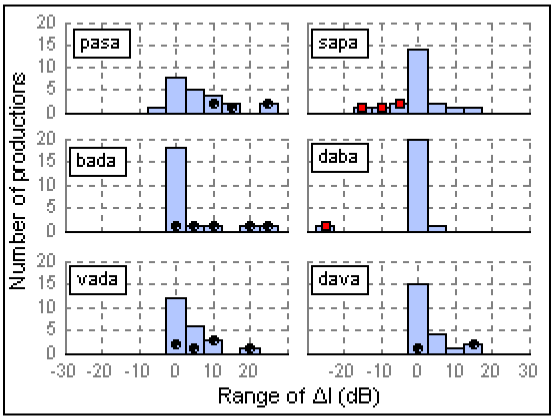

*Figure 4: Histograms of ΔI means across the 22 analyzed subjects, for each sequence. Points and square respectively display the number of positive and negative means significantly different from zero (t-test with Bonferroni correction, significant at the 0.05 level: p≤0.05/(22*6)).*

Vada/dava

The situation is different for /vada/ and /dava/, which share the same attractor: /vadá/→/vdá/ (Fig. 3., right). ΔI means are positive and significantly different from zero (4.0dB, p=0.002 for /vada/ and 3.2dB, p=0.006 for /dava/). Individual repartitions clearly confirm this trend (Fig. 4, third row).

4. Discussion

Globally, the results show a greater stability for LC CVCV sequences compared to CL ones. While LC utterances almost never switch towards the inverse pattern, switches from CL sequences to the LC symmetric pattern occur frequently, and even seem to be the rule in the case of the /dava/ CVCV. However, a number of differences between subjects and sequences are superimposed to this portrait.

4.1. Variability between subjects and between sequences

The results vary a lot between subjects, as displayed in Fig.4. That suggests that the dynamic system at work in the present experiment, consisting in the coordination between jaw, tongue and lips, is bi-stable. It may evolve towards two attractors, LCá and CLá. The global preference for LCá over CLá when rate increases could be modeled in the framework of synergetics and nonlinear oscillator theory, as done previously for other bi-stable human dynamic systems [11]. The differences between subjects could either be due to individual differences in dynamics and attractors, or to various initial conditions leading to fall in either one or the other attractor, though more frequently in the LCá attractor.

The differences between the /dava/ sequence and the two other CL sequences seem to indicate that the balance between attractors is different in this case, with a larger benefit for LCá over CLá. This could be due to articulatory differences between the labiodental /v/ and the bilabial /b/. For /v/, the jaw could stay high while for /b/, some downwards jaw movement is necessary for opening the mouth and producing the bilabial release. This could reinforce the possibility to synchronize the labial and coronal consonants in /vada/ and /dava/ sequences, better than in the four other sequences involving the bilabials /p/ or /b/.

4.2. A substance-based influence on phonological forms

The results of these experiments are in favor of the idea that LC sequences appear as a natural and coherent production unit, more stable and better "in-phase" than their CL inverse counterpart. This could explain the preference for LC utterances in infant production, rather than the "simple first" explanation previously proposed. There is actually no reason to believe that the greater stability of LC sequences in these experiments on adults could be due to a "simple first" mechanism. Rather, articulatory coherence could provide a common basis to both infant data and our results. It could also explain the asymmetry in verbal transformations [5], considering the role of perceptuo-motor interactions in speech perception in general [12] and in verbal transformations in particular [13].

The possible articulatory preference for LC structures could explain why languages combine syllables in the LC mode much more frequently than in the inverse order. Evidences for such kind of substance-based influence on phonological patterns are always difficult to obtain. It can generally be argued that phonology is the determinant of substance and not the inverse. In the present case, this would result in arguing that since LC structures are predominant in languages, they provide the target for speech production. However, there are two difficulties in this inverse reasoning. Firstly, the speeded reiterant paradigm is unlikely to suffer from much phonological influences, since it is rather conceived as a pure motor automatism. Secondly, the fact that the preference for LC structures is maintained for all studied sequences is particularly interesting, considering that phonological data are on the contrary quite different from one sequence to the other.

	LaCa/CaL	LaC_/CaL	LVCV	L/C
PS	5/1	59/16	138/120	1283/1371
BD	0/0	4/3	28/65	752/1275
VD	0/0	3/2	23/59	520/1275

Table 1: Relative frequencies of LC and CL associations in French (from Bdlex).

Table 1 presents the frequencies of existence of LaCa vs. CaLa, LaCV vs. CaLV and LVCV vs. CVLV syllables (V being any vowel), or words beginning by L or C, in French, according to the BdLex database [14]. The "labial first" pattern is very strong for /p/ compared with /s/, but it is more or less the inverse for /b/ and /v/ compared with /d/. This fact makes the predominant trend for both /vada/ and /dava/ to evolve towards the LC structure particularly interesting and significant in the present set of experiments. Note that differences between /p/ vs. /s/ and /v/ or /b/ vs. /d/ observed in table 1 could be explained by non-articulatory factors such as higher audibility of /d/ at the initial of an utterance compare to /v/ audibility. That could agree with the general assumption according to which languages would select phonologic forms that minimizing energetic cost while maximizing perceptive distinctiveness [15].

5. Conclusion

The LC pattern is more stable than the CL one, appearing as an attractor for the production of LC and CL sequences repeated at increasing rates. This provides a possible basis for understanding the LC effect displayed in phonology and development. The next step consists in finding what are the articulatory correlates of the present behavior, and how they can explain the better LC stability. This is the topic of a companion paper [16].

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 M. Sato for the data acquisition, I. Rousset and N. Vallée for their "BdLex assistance".

7. References

[1] Rousset, I. "From lexical to syllabic organization". *Proc. XVth ICPhS*, Barcelona, pp. 715-718, 2003.
[2] MacNeilage, P.F. and Davis, B.L. "On the origins of intenal 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] Vilain, A., Abry, C., Badin, P. and Brosda, S. "From idiosyncratic pure frame to variegated babbling : evidence from articulatory modelling", ICPhS'99, San Fransisco, USA. 1999.
[5] 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.
[6] Warren, M.R. Illusory changes of distinct speech upon repetition – The verbal transformation effect. British Journal of *Psychology*, 52: 249-258, 1961.
[7] Kelso, J.A.S., Saltzman, E.L., and Tuller, B. "The dynamical perspective on speech production: Data and theory", *J. of Phonetics*, 14, 29-59 and 171-196, 1986.
[8] Stetson, R.H. *Motor Phonetics: A study of speech movements in action.* Amsterdam: North-Holland, 1951.
[9] de Jong, K.J. "Rate-induced resyllabification revisited", *Language and Speech*, 44: 197-216, 2001.
[10] http://www.fon.hum.uva.nl/praat/
[11] Haken, H., Kelso, J.A.S., & Bunz, H. "A theoretical model of phase transitions in human hand movements", *Biological Cybernetics*. 51:347-356, 1985.
[12] Schwartz, J.-L., .Abry, C., Boë, L.-J. and Cathiard, M.A. "Phonology in a theory of perception-for-action-control", In: Durand, J., Lacks, B. (Eds.), *Phonology: From Phonetics to Cognition*. OUP, Oxford, pp. 240-280, 2002.
[13] Sato, M., Baciu, M., et al. "Multistable representation of speech forms: An fMRI study of verbal transformations", NeuroImage, 23, 1143-1151, 2004.
[14] http://www.elda.org/catalogue/fr/speech/S0004.html
[15] Liljencrants, J. and Lindblom, B. "Numerical simulations of vowel quality systems: the role of perceptual contrast", *Language*, 48:839-862, 1972.
[16] Rochet-Capellan, A., Schwartz, J.-L., "The LC effect and CVCV stability during reiterant speech production: An articulatory analysis", this conference submitted.