A study of anticipatory coarticulation for French speakers and for Mandarin Chinese speakers.

Liang Ma, Pascal Perrier, Jianwu Dang

To appear in Chinese Journal of Phonetics, 2, 2009,
http://www.phonetics.org.cn/phonetic/news/listnews.jsp?newsId=111

A study of anticipatory coarticulation for French speakers and for Mandarin Chinese speakers

Liang Ma, Pascal Perrier, Jianwu Dang

Abstract: Anticipatory coarticulation is studied for two languages, French and Mandarin Chinese, within $Vowel_1$ -Consonant -$Vowel_2$ sequences (V_1CV_2 henceforth). EMMA data and acoustic signals were collected. The influences of V_2 on V_1 and of V_2 on C are more specifically analyzed in this paper. The results showed that, for the French speakers, vowel V_2 influenced the whole sequence V_1CV_2, while its influence was limited to the syllable CV_2 for the Chinese speakers. This suggested that speech planning in French is managed beyond the size of a syllable, while the planning is limited within the syllabe for Chinese. The results demonstrated the impact of language based constraints on articulatory planning in speech.

Key words: Speech production, Coarticulation, Syllable, Speech planning

1. Introduction

Coarticulation is classically considered to be the consequence of two main factors: planning strategies of articulatory movement (cf. [5], [7], [13]) and some physical influences. Planning of speech production is more specific than that of other skilled human movements since it is constrained by linguistic factors linked to the phonological structure of the language. Our study focuses on the effects of the phonological structure on speech planning. To do so, we collected consistent kinds of utterances of French and Chinese, and then we analyzed anticipatory coarticulation using the VCV utterances.

2 Some phonological descriptions in Chinese and in French

2.1 Vowel inventories in Chinese and in French

The description of the vowel system of Chinese is quite controversial. According to a classical phonological description, there are five vowels, /i, y, u, a, ə/ in Mandarin Chinese [3], [6]. However, the situation is more complex in real phonetic environments, where a number of allophones emerge in the variant phonetic context. The vowels' variations are associated with 4 places of articulation (front, mid-front, mid-back and back), three degrees of aperture (high, mid, low) and the lip rounding/spreading characteristics.

French vowel inventory is described as including 11 oral vowels and four nasal vowels [2]. Distinction among the 11 oral vowels are realized in a three dimensional space, including three places of articulation along the front/back direction, four degrees of aperture in the vocal tract and the rounded/spreading characteristics of the lips for front vowels.

2.2 Status of the syllable in Chinese and in French

In Chinese the syllable is considered as the most important phonological unit. This is a very strong characteristic of the language. Firstly, as a written language, Chinese character corresponds to a syllable systematically. Secondly, as a tone language, each tone always concerns a syllable in its whole and it does not extent across syllable boundaries, although they may affect each other. Thus, the syllable specifies the positioning and the duration of the tones and it appears, thus, as a fundamental element structuring the language. For example, contrary to English or to French, which count many thousands of syllables, Chinese has only 400 monosyllables when tones are not taken in consideration [6]. Thus, syllable in Chinese appears to obey very specific rules and to exist in the linguistic structure of the language as a whole.

The status of syllable in French is different than that in Chinese. It is illustrated by the differences between the intonation systems of these two languages. Intonation in French is characterized by rising pitch movements occuring at phrase boundaries. The location of stress is fixed at the word level. A unit of intonation in French contains an average of 3.5~3.9 syllables e.g. [8]. Classically French is also considered as a syllable-based language since syllables seem to play a crucial role in speech segmentation [11]. However some recent experiments have softened this hypothesis of the dominance of syllables at the level of the phonological representation of French. Indeed, Content [4] suggested that the syllable effect in speech segmentation in French could be, at first, due to specific acoustic–phonetic properties existing in the acoustic signal at syllables boundaries and not to a top-down process involving syllabic representations. Although it is not denied that syllable is a fundamental element in French, it can be speculated that the strength of syllable at the phonological level is weaker than it was originally supposed to be, and less dominant than that in Chinese.

3 Speech material and Data analysis

Speech material consists of 15 VCV nonsense words where the vowel was /a/, /u/ or /i/ and the consonant was /k/ or /t/. The words were uttered at a normal speech rate by three native French speakers (AV, PB, CV) and two native Chinese speakers (SK and JW). Each target word was embedded in a carrier sentence: "C’est VCV ça?" in French and "这是VCV 吗?" in Chinese. Each carrier sentence was repeated 10 times. In order to have consistent sequences in both languages, influence of the tonal structure for Chinese sequences was avoided. All the sequences in Chinese were produced with a high level tone. We also did not include sequences such as /aki/, /iki/ and /uki/ that do not exist in Chinese.

The articulatory data were collected with an electromagnetic midsagittal articulograph (EMMA; AG100 Carstens Electronics). Four sensors glued on the tongue are called T_1, T_2, T_3 and T_4, from the tongue tip to the tongue back. One sensor was also glued on the upper lip, one on the lower lip, and one on the lower incisor. Two reference sensors were located on the upper incisor and on the bridge of the nose.

In a first step, the labeling was carried out manually on the spectrogram. For the consonants, the onset of the burst was measured. For the vowels, the point with maximum stability of the first three formants on the spectrogram was labeled. Then, in a second step, in order to achieve a more accurate detection of the most canonical vocal tract configuration, the label for vowels was automatically moved towards the extreme position of the tongue back sensor T_4 in the mid sagittal plane.

For each sequence V_1CV_2 and for each subject, the influences of V_2 on the articulation of the preceding phonemes were statistically assessed. A variance analysis ANOVA (Repeated Measures) and a post-hoc test were carried out. SPSS for windows was used for this analysis. It is important to mention that the accuracy of the EMMA system AG100 was estimated to be around 0.5 mm. For this reason, we considered differences in tongue positioning to be physically significant only if they were larger than 0.5 mm. In this case, statistical significance threshold was at $p<0.05$.

Because of hardware problems that happened during the experiment, a large amount of data was missing on sensor T_3 for subject AV and on sensor T_1 for subjects CV and SK. Hence, these sensors were not taken in consideration in the analysis.

4 Results

4.1 Reference tongue shapes for vowels /a, u, i/ in Chinese and in French

Figure 1 shows the tongue sensors positions for each vowel V_1 in the symmetrical /V_1tV_2/ sequence, in which $V_1=V_2$, for each subject. The dispersion ellipses were calculated from the measurements of all repetitions of sequences. The apical consonantal context was chosen in order to minimize the possible impact of the consonant on the vowel tongue shape. The symmetrical context was selected to avoid any form of anticipatory vowel-to-vowel coarticulation. These articulatory configurations can be considered as the reference patterns for the 3 vowels /a/, /u/ and /i/ in French and in Chinese. Although certain differences exist among speakers, the common features can be extracted as follow: For vowel /a/, the tongue shape is essentially flat for all the subjects. For vowel /u/, sensor T_4 is located in the region of articulation for all the subjects. For vowel /i/, sensors T_2 and T_3 are close to the palate and they correspond for all the subjects to the region of articulation.

These characteristics served as reference to analyze the whole set of data and to interpret them in terms of anticipatory coarticulation. It can be expected that, if the production of V_2 is anticipated during the production of the preceding phonemes (V_1 and C), the sensors positions of these preceding phonemes will be influenced as follows: (1) If the coming phoneme V_2 is vowel /a/, T_3 and T_4 of current phoneme should be lower than if V_2 is vowel /i/ or vowel /u/. (2) If V_2 is vowel /u/, T_4 of current phoneme should be higher and more posterior than if V_2 is vowel /i/ or vowel /a/. (3) If V_2 is vowel /i/; T_2 and T_3 should be higher and more anterior than if V_2 is vowel /a/ or vowel /u/.

4.2 Effects of V_2 on V_1

Figure 2 shows tongue sensors positions for V_1 in the various V_2 contexts in /V_1tV_2/sequences for each subject (in rows) and for each vowel V_1 (in columns). The tongue position of vowel V_1 was characterized only with the three sensors T_2, T_3 and T_4, when they were available. Since the tongue tip is underspecified for vowels, T_1 was not considered.

The mean differences of tongue positions for V_1 associated with the changes in V_2 are shown for /V_1tV_2/ sequences in Table 1. Only the differences that are proved to be statistically significant ($p<0.05$) by a post-hoc test carried out after the ANOVA analysis are presented in these tables. (Table 1.a for three French speakers and table 1.b for two Chinese speakers). At the first sight, one can see that there are much more cases with significant differences of V_1 for the French speakers than for the Chinese speakers. Looking more precisely on the directions of the differences, it can be noted that for the French speakers the significant differences of V_1 are all compatible with the predictions of an anticipatory coarticulation of V_2, which are listed above, except in three unexpected cases (grey shaded in Table 1.a) observed for subject AV when V_1 was vowel /i/. On the contrary, for the Chinese speakers, among the few significant differences that are shown in Table 1.b, less than the half is compatible with predictions based on anticipatory coarticulation.

These results suggest that for the French speakers, vowel V_2 influences the tongue shape during the production of V_1, while, for the Chinese speakers, this influence of V_2 on V_1 is not only very small, but also the indirect result of an anticipatory strategy. Similar results are obtained for the palatal consonant context sequences /V_1kV_2/.

4.3 Effects of V_2 on C

Figure 3 shows tongue sensors positions measured for consonant /t/ in the various V_2 contexts for the 5 subjects and for the different vowels V_1. One can see that the scatters of T_2, T_3 and T_4 show larger influences of V_2. The clearest and the most robust influences are in the vertical direction for both French speakers and Chinese speakers, while some variability exists also in the horizontal direction.

The average differences in tongue positions for /t/ associated with the changes in V_2 were measured using the same method as for vowel V_1. For both French speakers and Chinese speakers, we have observed that, in the vertical directions, the significant average differences of T_2, T_3 and T_4 related to V_2 were numerous and they were almost all in agreement with the predictions of anticipatory strategy. But in the horizontal direction, numbers of the influences were in contradiction with our predictions for both two languages. This can be explained from the point of view of the constraints associated with the production of the alveolar stop consonant, which requires an occlusion in the very front part of the vocal tract. Such a constraint is very crucial and allows little freedom for change by the anticipation.

For the palatal consonnant /k/ sequences, significant influences of the vowel V_2 on the consonant C were also observed for both French speakers and Chinese speakers.

5 Discussion and conclusions

Our results showed that, for the French speakers, vowel V_2 influences the whole sequence V_1CV_2, while its influence is limited to the syllable CV_2 for the Chinese speakers. These findings suggest that French speakers and Chinese speakers use different strategies on speech planning.

A possible explanation for the differences in the anticipation of V_2 in V_1 could lie in the vowel inventories of these two languages. Indeed, density of vowel inventory has been suggested to influence vowel variability associated with vowel to vowel coarticulation [10]: the larger the vowel inventory, the smaller the variability. However, in spite of the fact that the description of the vowel system of Chinese is still a controversial issue, it is possible to say that it has fewer vowels than French. From this perspective, the variability of V_1 in Chinese should be larger or equal to that of

V_1 in French. But we observed the contrary.

We rather interpret the differences between the anticipatory coarticulation strategies of the French and those of the Chinese speakers as the consequences of the respective status of the syllable in their languages. As mentioned in the phonological description for these two languages, the influences of the CV syllable are stronger in Chinese than in French. For Chinese, the syllable is dominant in linguistic representation. Chinese speakers seem to use the syllable as a basic unit to produce sequence. This explanation is compatible with the syllable model of coarticulation [9] about the major role of the syllable in the organization of speech. On the other hand, French speakers seem to use other longer sequence than syllable on speech planning. Our results for French speakers confirm models of coarticulation like Öhman's model [12], the MEM model [1], which all take into account phoneme sequences longer than the CV syllable.

Further work using different models of control applied to a biomechanical model of the tongue will aim at testing these different hypotheses.

6. References

[1] Abry, C., Lallouache, T. M. (1995) Modelling lip constriction anticipatory behavior for rounding in French with the MEM (Movement expansion Model). *Proceedings of the 13th International Congress of Phonetics Sciences, 4*, 152-155.

[2] Calliope (1989). *La parole et son traitement automatique.* Paris: Masson

[3] Chao, Y. R. (1968) *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

[4] Content, A., Meunier, C., Kearns, R. K., Frauenfelder, U. H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes, 16,* 609-636.

[5] Dang, J., Akagi, M., Honda, K. (2006) Communication between speech production and speech perception within the brain- Observation and simulation, *Journal of computer science and technology, 21, 1,* 95-105.

[6] Deng, L., Dang, J. (2007) "Speech Analysis: The Production-Perception Perspective," in Hai-Zhou Li and Chin-Hui Lee (eds.), *Advances in Chinese Spoken Language Processing,* 3-32, Publisher: World Scientific, New Jersey, 2007.

[7] Jordan, M. I. (1989) Serial order: A parallel distributed processing approach. In J. L. Elman and D. E. Rumelhart (Eds.), *Advances in Connectionist Theory: Speech*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[8] Jun, S.A., Fougeron, C. (2002) The Realizations of the Accentual Phrase in French Intonation, *Pobus 14: a special issue on Intonation in the Romance Languages*, 147-172, J. Hualde (Ed.).

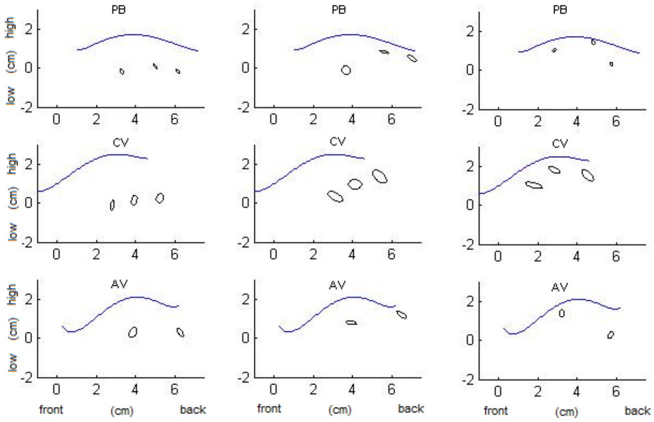
[9] Kozhevnikov, V., Chistovich, L. (1965) *Speech: Articulation and Perception*, 104-118, Washington DC: Joint Publication Research Service.

[10] Manuel, S. (1990), The role of contrast in limiting vowel-to-vowel coarticulation in different languages, *Journal of the Acoustical Society of America 88,* 1286-1298.

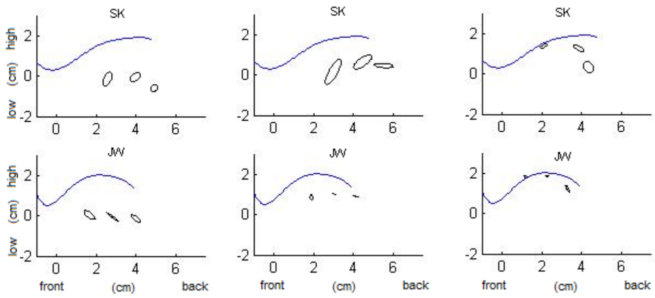
[11] Mehler, J., Dommergues, J.-Y., Frauenfelder, U., Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298–305.

[12] Öhman, S. (1966) Coarticulation in VCV utterance: Spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 151-168.

[13] Whalen, D. H. (1990) Coarticulation is largely planned. *Journal of Phonetics 18,* 3-3.

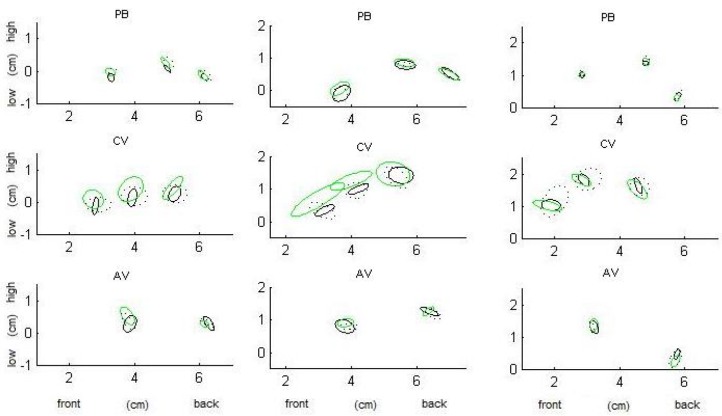


(a) French speakers

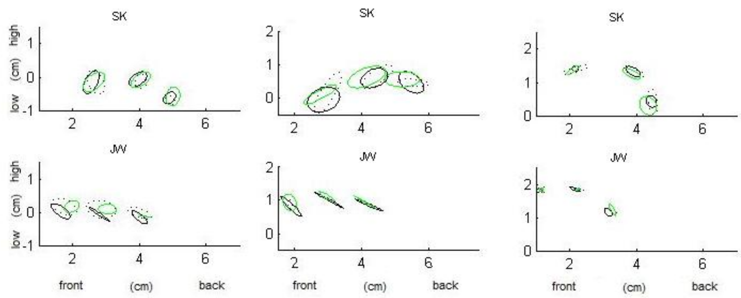


(b) Chinese speakers

Figure 1: A general view of tongue sensors positions (cm) for three vowels V_1 /a, i, u/ in symmetrical /V_1tV_2/sequences with the palate contour as a geometrical reference for 3 French speakers (a) and two Chinese speakers (b) . Sensors T_2, T_3 and T_4 (from left to right) are presented when they are available. Left panel: V_1=/a/ in /ata/ sequences, middle panel V_1=/u/ in /utu/ sequences; right panel: V_1=/i/ in /iti/ sequences;

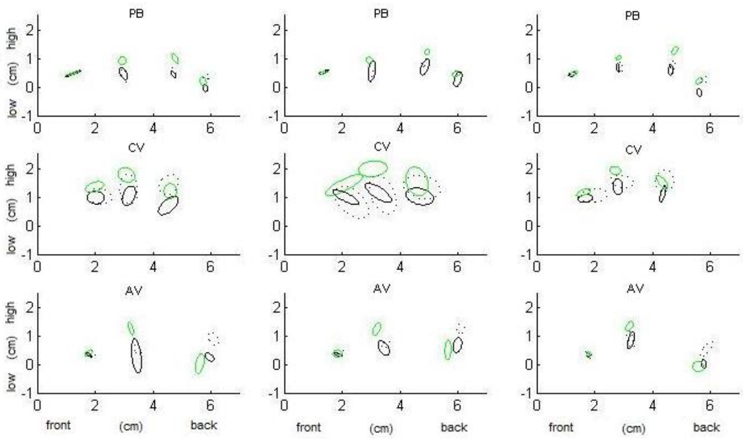


(a) French speakers

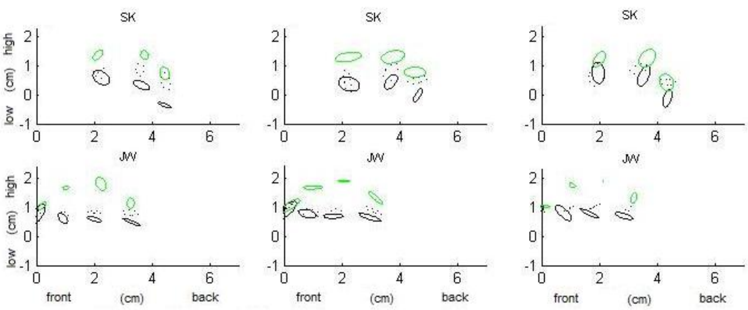


(b) Chinese speakers

Figure 2: Tongue sensors positions (cm) for 3 vowels V_1 in $/V_1tV_2/$ sequences for 3 French speakers (a) and two Chinese speakers (b). Sensors T_2, T_3 and T_4 are presented. Left panel: V_1=/a/; middle panel: V_1=/u/; right panel: V_1=/i/. Black line: V_2=/a/; Gray line: V_2=/i/; dotted line: V_2=/u/.



(a) French speakers



(b) Chinese speakers

Figure 3: Tongue sensors positions (cm) for C= /t/ in $/V_1tV_2/$ sequences for 3 French speakers (a) and two Chinese speakers (b). Sensors T_1, T_2, T_3 and T_4 are presented when they are available. Left panel: V_1=/a/; middle panel: V_1=/u/; right panel: V_1=/i/. Black line: V_2=/a/; Gray line: V_2=/i/; dotted line: V_2=/u/.

Table 1. Mean differences of tongue sensors positions (cm) for V_1 associated with the change of V_2 in /V_1tV_2/ sequences. Only the statistically significant cases are presented (*p<0.05*) and only if the difference amplitude is larger than 0.5 mm. Positive differences in the X-axis indicates that the sensor is more posterior, and for the Y-axis it means the sensor is higher. The black shades show the cases that the sensor is not available and the grey shade for the cases that the observation is opposite to the prediction of anticipation.

(a). Three French speakers PB, CV and AV

	PB			CV			AV		
/a/	/atu/-/ata/	/ati/-/ata/	/atu/-/ati/	/atu/-/ata/	/ati/-/ata/	/atu/-/ati/	/atu/-/ata/	/ati/-/ata/	/atu/-/ati/
T_{2x}			0.08					-0.08	
T_{2y}	0.19	0.16			0.21	-0.17		0.24	-0.16
T_{3x}			0.09						
T_{3y}	0.16	0.18			0.26	-0.20			
T_{4x}		-0.06	0.08					-0.13	0.07
T_{4y}					0.19	-0.18			
/u/	/utu/-/uta/	/uti/-/uta/	/utu/-/uti/	/utu/-/uta/	/uti/-/uta/	/utu/-/uti/	/utu/-/uta/	/uti/-/uta/	/utu/-/uti/
T_{2x}							0.14		0,11
T_{2y}			-0.18		0.33	-0.34		0.13	-0.15
T_{3x}									
T_{3y}		0.07			0.26	-0.30			
T_{4x}					-0.24				0.15
/i/	/itu/-/ita/	/iti/-/ita/	/itu/-/iti/	/itu/-/ita/	/iti/-/ita/	/itu/-/iti/	/itu/-/ita/	/iti/ -/ita/	/itu/-/iti/
T_{2x}						0.22	-0.06		-0.07
T_{3x}				0.15		0.22			
T_{3y}	0.08								
T_{4y}			0.10				-0.14	-0.20	

(b). two Chinese speakers SK and JW

	SK			JW		
/a /	/atu/-/ata/	/ati/-/ata/	/atu/-/ati/	/atu/-/ata/	/ati/-/ata/	/atu/-/ati/
T_{2x}					0.32	-0.23
T_{3x}					0.24	-0.20
T_{3y}				0.16	0.16	
T_{4x}		0.11			0.17	-0.15
T_{4y}				0.12		
/i /	/itu/-/ita/	/iti/-/ita/	/itu/-/iti/	/itu/-/ita/	/iti/-/ita/	/itu/-/iti/
T_{4x}		-0.09	0.14	0.14	0.13	