AN ACOUSTIC STUDY OF V-TO-V COORDINATION WITHIN AND ACROSS WORDS IN FRENCH

Alice Yildiz, Daria D'Alessandro, Sejin Oh, Anne Hermes, Cécile Fougeron

Laboratoire de Phonétique et Phonologie, UMR7018, CNRS/Sorbonne Nouvelle (France)
alice.yildiz@sorbonne-nouvelle.fr, daria.dalessandro@sorbonne-nouvelle.fr, se-jin.oh@cnrs.fr, anne.hermes@sorbonne-nouvelle.fr, cecile.fougeron@sorbonne-nouvelle.fr

ABSTRACT

To better understand coordination beyond the syllable, we investigate inter-syllable coordination in French, depending on the boundary separating the syllables, i.e., a syllable boundary within a word vs. a word boundary in a subject-verb phrase. Patterns of coordination are inferred by acoustical measures of V-to-V anticipatory coarticulation and temporal V-to-V lag. The effect of the intervening consonant (/p/ vs. /ʁ/), as well as the variability in the spectral and lag characteristics across 45 repetitions, is compared across the two boundary types. Coarticulation decreases across-words compared to within-word, but only with the consonant /p/, showing the least articulatory constraint, while token-to-token variability is similar in the two boundary contexts. V-to-V lag is longer across- than within-word. Interestingly, variability in lag duration is found to increase across-words, but only with /p/. Overall, within-word cohesion between syllables translates in a specific pattern of coordination only in the context favoring gestural overlap.

Keywords: anticipatory V-to-V coarticulation, V-to-V duration, inter-syllabic coordination, consonantal constraint, variability

1. INTRODUCTION

Even though a huge amount of work has been done on intergestural coordination, several aspects underlying this process remain poorly understood. One unresolved matter relates to the speech units over which speech gestures are coordinated with each other. Indeed, if a special cohesion is considered to characterize tautosyllabic gestures, more unclear is how speech gestures are organized over units larger than the syllable, especially within a word. In particular, the question is open whether the segments composing a lexical unit present a specific coordination pattern which reflects within-word gestural cohesion; in other terms, whether a lexical word is a unit of gestural organization. Studies that have compared intergestural coordination within and across words by looking at CV coarticulation or overlap in CC sequences reported no effect of word boundary on the degree of overlap [1, 2, 3], or inconsistent results [4, 5, 6]. However, some studies looking at acoustic rhyme lengthening reported longer rhymes at word boundaries than in word-internal position [7, 8]. In a first investigation of anticipatory V-to-V coarticulation across syllables within vs. across words in French, [9] found that four out of five speakers coarticulated more within a word than across words, even in cases where the two words belonged to the same accentual phrase (i.e., they were not separated by a boundary other than a lexical or prosodic word boundary).

In this acoustic study, we tackle further the specificity of word-internal coordination. Thus, we infer coordination between syllables within- and across-words indirectly by spectral cues of anticipatory V-to-V coarticulation, but also by the temporal acoustic lag between the two vowels. Following the assumption that intergestural coordination in a unit where the coordination is planned would result in more stable patterns [10], we test whether spectral cues of coarticulation and temporal lag are more stable within-words, when looking at multiple repetitions of the same tokens by five speakers. Moreover, following the DAC model [11], according to which constraints on the articulation of the consonants should affect V-to-V lingual anticipation, we test our acoustic proxies of coordination in sequences favoring overlap (/VpV/) or disfavoring (/VʁV/) it.

2. METHOD

2.1. Speech material and procedure

Target sequences were /pV_1.CV_2/ disyllabic items where V_1 was /a/, C_2 was either /p/ or /ʁ/ and V_2 was /i/. The boundary between the two syllables was either a syllable boundary within a word (within-word position, e.g., in P<u>aris</u>) or a word boundary between a subject and a verb (across-word position, e.g., in *pap<u>a rit</u> beaucoup*, Engl. "Dad laughs a lot"). Target sequences were embedded in meaningful sentences (see Table 1) and recorded in five subsequent sessions. Each session contained the sentences in a pseudo-randomized order, with each target sequence occurring nine times per session, leading to a total of

45 repetitions per speaker. This procedure was followed to collect a large number of repetitions and to test the variability in the production of the target sequences. Five female native French speakers, aged 22 to 28, participated in the recordings.

Sequence	Sentence (example)
/a.pi/ within-word	Quand p**api** l'aura vu, il nous croira. *When Grandpa sees him/her/it, he'll believe us.*
/a#pi/ across-words	Pap**a pi**lote un avion de ligne. *Daddy flies an airliner.*
/a.ʁi/ within-word	Quand P**aris** est sous la neige, c'est très beau. *When Paris is under the snow, it is very beautiful.*
/a#ʁi/ across-words	Pap**a ri**t beaucoup au cinéma. *Daddy laughs a lot at the movies*

Table 1: Sentences per boundary and consonant types, for the V_2 = /i/ context.

2.2. Measures

A first proxy of coordination was defined as the spectral influence of V_2 /i/ on V_1 /a/ in /api/ and /aʁi/ sequences, with /apa/ and /aʁa/ sequences being used as a control context. Since the anticipation of /i/ on /a/ should translate into a lowering of F1 and an increase in F2, a composite measure of F2-F1 compactity (in Hz), was taken for V_1 /a/ and V_2 /i/. This F2-F1 compacity was used to carry out two analyses:
(a) first, to assess whether there is anticipatory V-to-V coarticulation for both boundary conditions and intervening consonants, F2-F1 compacity of V_1 /a/ followed by V_2 /i/ is compared to the one of V1 /a/ followed by V_2 /a/. A higher compacity is expected for V_1 /a/ when V_2 is /i/.
(b) second, to compare conditions in terms of the degree of coarticulation, a measure of acoustic assimilation of V_1 /a/ to V_2 /i/ is computed within each /aCi/ token as the difference between the compacity of /a/ and /i/ over the compacity of /i/:

(1) *acoustic assimilation index:*

$$\frac{((F2-F1)_{/a/}-(F2-F1)_{/i/}}{(F2-F1)_{/i/}}$$

A higher acoustic assimilation index corresponds to a higher degree of coarticulation.

The second proxy of coordination relies only on temporal information. V-to-V acoustic lag is measured as the distance from the acoustic onset (periodicity in the signal and apparition of formants) of V_1 to the acoustic onset of V_2. This lag covers V_1.C duration and a possible break between the two syllables, even if a pause is not expected within a word, or between a subject and a verb in the across-word condition.

For both the acoustic assimilation index and the V-to-V lag measure, variability across repetitions is measured as the absolute deviation of each token from the grand mean in each condition per speaker.

2.3. Statistical analysis

The data were analyzed with either linear mixed models, used for the analyses on coarticulation, or generalized mixed models, used for the analyses on V-to-V lags and variability. The choice between a linear or generalized model was motivated by data distribution. Models were built in the R environment with the lmer() or glmer() functions of the lme4 package [12]. The best-fitting distribution of the generalized linear mixed model was chosen by comparing AIC values. A gamma distribution was selected for the analyses on variability (log link for acoustic assimilation variability and inverse link for V-to-V lag duration variability) while an inverse Gaussian ($1/\mu^2$ link) distribution was chosen for V-to-V lag.

To test for the presence of coarticulation, intended as a spectral difference between /aCa/ and /aCi/ sequences, a model was run with F2-F1 compacity of V_1 /a/ as dependent variable and BOUNDARY ('within-word' vs. 'across-words'), CONSONANT (/p/ vs. /ʁ/) and V_2 TYPE (/a/ vs. /i/) as fixed effects and their interaction.

To compare conditions and C articulatory constraints, four models were run with *acoustic assimilation index*, *assimilation index variability*, *V-to-V lag*, and *lag variability* as dependent variables and BOUNDARY and CONSONANT as fixed effects and their interaction. Random intercepts for SPEAKER and REPETITION were included (except for the models on variability).

The effect of each fixed factor and interaction on the dependent variable was tested either by comparing the model with the given interaction/factor with a model lacking that particular interaction/factor by performing a chi-squared test, or with a simple type III ANOVA analysis. Post-hoc comparisons were performed with the *emmeans* package [13].

3. RESULTS

3.1. F2-F1 compacity

Presence of V_2 to V_1 anticipatory coarticulation is found regardless of consonant or position, as shown by a higher F2-F1 compacity of V_1 /a/ when it is followed by V_2 /i/ than by V_2 /a/ (as illustrated by the pairwise comparisons in Table 2). However, the magnitude of the difference between the two vocalic contexts depends on consonant type and position (see Table 3).

BOUNDARY	V_2	mean (sd)	*p-value*
/p/			
within-word	/a/ /i/	870 (129) 1131 (171)	$p <.001$
across-words	/a/ /i/	912 (136) 1028 (142)	$p <.001$
/ʁ/			
within-word	/a/ /i/	673 (164) 869 (140)	$p <.001$
across-words	/a/ /i/	788 (137) 878 (156)	$p <.001$

Table 2: F1-F2 compacity of V_1 /a/ according to V_2 for each boundary and consonant condition.

Fixed effects	*p*-value
V_2 TYPE	$p <.001$
V_2 TYPE*BOUNDARY	$p <.001$
V_2 TYPE*CONSONANT	$p <.001$
V_2 TYPE*BOUNDARY* CONSONANT	$p =.03$

Table 3: Type III ANOVA analysis results on the relevant fixed effects and interactions of the model for F2-F1 compacity measure.

3.2. Acoustic assimilation index

As illustrated in Fig. 1, a higher degree of coarticulation, corresponding to higher values of the acoustic assimilation index, is found within-words than across-words, but only in the /p/ context ($p < .001$) and not in the /ʁ/ context. A summary of the fixed effects and their interaction is given in Table 4.

Fixed effects	*p*-value
BOUNDARY	$p <.001$
BOUNDARY*CONSONANT	$p <.001$

Table 4: Type III ANOVA analysis results on the model's relevant fixed effects and interactions for acoustic assimilation index.

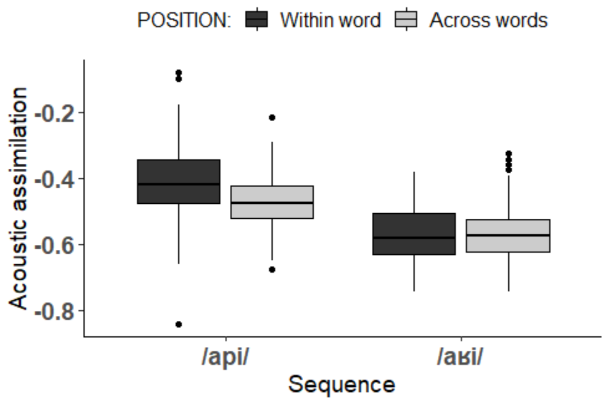

Figure 1: Acoustic assimilation index of the V_2 /i/ on the V_1 /a/ across consonant type (/p/ and /ʁ/) in the two boundary conditions (within-word and across-words). The higher the index, the more there is coarticulation.

3.3. V-to-V lag

V-to-V lag is significantly longer across-words than within-word ($z = 7.571, p < .001$) in both consonantal context: there is no interaction between BOUNDARY and CONSONANT ($X^2(1) = 0.46, p = n.s.$).

3.4. Variability analysis

The variability in acoustic assimilation index does not depend on the boundary condition or consonant type, as shown by the fact that none of the factors contribute significantly to the model.

However, a boundary effect is found for the V-to-V lag measure, but in interaction with the consonant type (BOUNDARY*CONSONANT $p < .001$). Indeed, as illustrated in Fig. 2, V-to-V lag duration is found to be more variable across-words than within-words only in the /p/ context ($z = 6.746, p < .001$), but not in the /ʁ/ context ($z = -1.113, p = n.s.$).

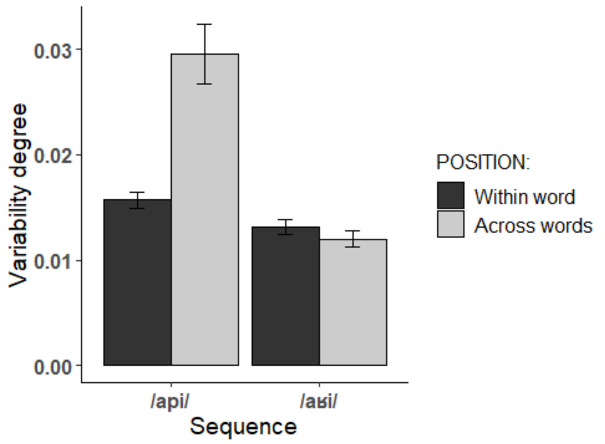

Figure 2: Variability in V-to-V lag (absolute deviation from the grand mean) across boundary conditions per consonant types.

4. DISCUSSION AND CONCLUSION

In this study, we investigated possible acoustic consequences of a difference in coordination between syllables within- and across-words, by looking at both spectral and temporal domains.

The first main result of this study is that there is an anticipation of V_2 in V_1, in terms of spectral assimilation, regardless of the type of intervening consonant and whether this anticipation is observed within or across words, in line with previous studies observing V-to-V coarticulation across one or more words, e.g. [14, 15]. More interestingly, this coarticulation is found to be larger within a word, but only when the consonant separating the two vowels is a labial consonant. This consonant-dependent effect of the boundary condition can be accounted for by the different articulatory constraints in the two types of consonantal context. With an intervening /p/ between the two vowels, the tongue is freer to vary and to adapt its trajectory according to V_2 height. Across /ʁ/, V-to-V coarticulation exists but is much reduced in terms of spectral effects on V_1 (smaller acoustic differences in compacity and lower acoustic assimilation index). The uvular articulation of /ʁ/ entails an adaptation of the V-to-V trajectory. The acoustic consequence of this articulatory constraint is that the contextual effects of /ʁ/ and /i/ on V_1 /a/ are antagonistic: /ʁ/ increases F1 and decreases F2 of the surrounding vowels [16], while /i/ pushes /a/ to a lower F1 and higher F2. With a reduced degree of freedom of V-to-V lingual anticipation in this context, it is possible that differences in coarticulation within- vs. across-words do not surface in the acoustic signal, or do not exist.

Nonetheless, the absence of difference in coarticulation variability shows that the degree of V_2 anticipation is equally stable regardless of the intervening consonant or boundary condition. This stability in the spectral cues of this V-to-V anticipation across repetition is in-line with [10], suggesting that V-to-V anticipation is controlled.

Indication of a higher intersyllabic cohesion within- than across-words comes also from the results on the temporal measures. Indeed, V-to-V lag is consistently longer across a word boundary regardless of the different articulatory constraints imposed by the intervening consonant. This is consistent with previous acoustic studies showing final lengthening at the word boundary (in a sequence composed of two content words, see [8]). Moreover, more variability is found in V-to-V lag duration across word boundaries, again in the freer /p/ context. A more stable timing within-words reinforce the picture of tighter gestural coordination within a lexical unit: for instance, Byrd [4] reported less variability in timing for syllable onset clusters, which are considered to have specific timing relationships, than for clusters spanning a word boundary. However, in this case, the difference within- vs. across-words is not found in /ʁ/ context, for which there is no difference in lag variability depending on position, similar to the pattern found for the measure of anticipatory V-to-V coarticulation. Again, the stronger constraint imposed by the lingual consonant and its coordination with the following vowel in the /ʁV_2/ syllable may hinder a boundary effect.

Taken together, our results provide evidence of a greater cohesion between two syllables belonging to the same lexical word, as suggested by a higher degree of acoustic assimilation (only in /p/ context) and longer lag between V_1 onset and V_2 onset (in both C contexts), and a more stable lag in the /p/ context. Follow-up studies, incorporating articulatory and acoustic data, are on the agenda to further investigate cues of a specific gestural organization within lexical units, as compared to various types of word boundaries in controlled prosodic constituents smaller than the accentual phrase in French.

5. ACKNOWLEDGEMENTS

This study was funded by ChaSpeePro (CRSII5_202228) of the Swiss National Science Foundation and by the program "Investissements d'Avenir" ANR-10-LABX-0083 (Labex EFL).

6. REFERENCES

[1] Daniloff, R., Moll, K. 1968. Coarticulation of Lip Rounding. *Journal of Speech and Hearing Research*, 11(4), 707-721.

[2] Moll, K. L., Daniloff, R. G. 1971. Investigation of the Timing of Velar Movements during Speech. *The Journal of the Acoustical Society of America*, 50(2B), 678-684.

[3] Lewis, J., Daniloff, R., Hammarberg, R. 1975. Apical coarticulation at juncture boundaries. *Journal of Phonetics*, *3*(1), 1–7.

[4] Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209-244.

[5] Hardcastle, W. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, *4*(1–3), 247–263.

[6] Hardcastle, W., Roach, P. 1979. An instrumental investigation of coarticulation in stop consonant sequences. *Current Issues in the Phonetic Sciences*, 531.

[7] Sugahara, M., Turk, A. 2004. Phonetic reflexes of morphological boundaries at a normal speech rate. *Speech prosody 2004*, 353–356.
[8] Turk, A., Shattuck-Hufnagel, S.I. 2000. Word-boundary-related duration patterns in English. *J. Phonetics*, 28, 397-440.
[9] D'Alessandro D., Yildiz A., Fougeron C. 2022. Variation individuelle de la coarticulation en fonction de la frontière prosodique. *XXXIVe Journées d'Études sur la Parole*, Noirmoutier, France.
[10] Whalen, D. H., Chen, W. R. 2019. Variability and Central Tendencies in Speech Production. *Frontiers in Communication*, 4.
[11] Recasens, D., Pallarès, M. D., Fontdevila, J. 1997. A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, *102*(1), 544-561.
[12] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
[13] Lenth, R. 2022. Estimated Marginal Means, aka Least-Squares Means. *R package version 1.7.3.*
[14] Abry C., Lallouache, T. 1995. Le MEM : Un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français. *Bulletin de la Communication Parlée*, 3, 85-99.
[15] Grosvald, M. 2009. Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, *37*(2), 173-188.
[16] Guitard-Ivent, F., Turco, G., Fougeron, C. 2021. Domain-initial effects on C-to-V and V-to-V coarticulation in French: A corpus-based study. *Journal of Phonetics*, *87*, 101057.