INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

The effects of prosody on French V-to-V coarticulation: A corpus-based study

Giuseppina Turco[1], *Cécile Fougeron*[1], *Nicolas Audibert*[1]

[1] Université Sorbonne-Nouvelle Paris 3 - Laboratoire de Phonétique et Phonologie (LPP), UMR 7018, 19 rue des Bernardins

giuseppina.turco; cecile.fougeron; nicolas.audibert@univ-paris3.fr

Abstract

This study examines whether the degree of vowel-to-vowel coarticulation in French (better known as "vowel height harmony", V2-to-V1 henceforth) varies as a function of position in prosodic domain (i.e. IP initial vs. word-medial) and duration of V1 (i.e. short vs. long). Following the literature on the phonetics-prosody interface, segments at stronger edges are more resistant to coarticulatory effects induced by their neighboring vowel. While previous studies have mainly looked at non-local V-to-V coarticulation across prosodic boundaries/domains (e.g.,V#(C)V), here we look at V2-to-V1 coarticulation within an Intonational Phrase according to whether target V1 is in absolute initial position (#V1C(C)V2, e.g., #*essaie* [esɛ]/[ɛsɛ] - 'try') or not (word-medial, e.g., *épaissit* [epɛsi]/[epesi] - 'thikened'). The analyses are based on 33k words presenting possible V1C(C)V2 harmonic contexts, which were extracted from a corpus of French running speech. V2-to-V1 coarticulation is measured as the lowering of the first formant of the target V1 (/e, ɛ, o, ɔ/) in relation to the height of the V2 trigger /+high/ (i.e. mid-high and high) vs. /-high/ (i.e. mid-low and low). Results show an effect of prosodic position (but no effect of V1 duration) on V2-to-V1 coarticulation: V1 is more resistant to coarticulation when initial in an IP.

Index Terms: vowel harmony, coarticulatory resistance, phonetics-prosody interface

1. Introduction

Degree of anticipatory V2-to-V1 coarticulation affecting vowel height in French has been often referred to as a phonological – though optional – process of vowel *harmony*. In the French word *aime* ("(I) love"), the first vowel is typically produced as mid-low [*ɛm*], compared to the first vowel of *aimer* ("to love"), which tends to be produced as mid-high [*eme*] due to the closing influence of the mid-closed final vowel. Vowels targeted by this process are the mid French vowels (/e, ɛ, œ, ø, o, ɔ/), which can alternate in height (mid-high/mid-low) according to the height of the following V2 trigger.

While the exact description of the phenomenon may vary in the classical French literature (cf. [1-3]), recent acoustic studies [4, 5] have provided empirical evidence for its sensitivity to several factors. So far, the following effects on V2-to-V1 coarticulation have been attested: a) *V1 backness*: coarticulation is stronger for back V1 than front V1 [4, 5]; b) b) *type of intervocalic segment*: it is stronger when the intervening consonant is represented by a labial than a lingual consonant and with consonant clusters compared to consonant sequences resulting from schwa drop [5]; c) *speaking variety*: it is more systematic in Northern French speaking varieties than in Southern French varieties [4]; d) *orthographic influence*: it is stronger when not biased by orthography [5]. In the current study, we further look at factors that may influence the V-to-V coarticulatory process in a large database of natural speech, by examining two other well-known sources of variation: position-in-prosodic-domain and vowel duration.

Within the wide range of studies falling under the rubric of *prosodic strengthening* (i.e. spatio-temporal expansion of segments) at prosodic domain edges (see [6] for a review), very little work has so far been devoted to test the idea that strengthening in prosodically strong position is related to coarticulatory resistance (e.g., [7, 8]). Seminal work on V-to-V coarticulation by [9] in English, followed by [10] in Italian, has shown that the amount of V-to-V coarticulation across a prosodic boundary can be affected by the prosodic strength of this boundary. Hence, in a CV1#CV2 sequence, the post-boundary vowel V2 tends to resist coarticulatory influence of the pre-boundary V1 when CV2 is initial in a prosodically strong prosodic domain. If these results bring support to the idea that initial position in a high prosodic domain is somehow a 'preserved' position, it remains unclear whether the reduced coarticulation on V2 is due to the fact that coarticulation is blocked by the presence of a strong boundary between the target and the trigger, or to the fact that the domain initial position is resistant *per se* (i.e. whether it is a "boundary" vs. "edge" effect).

Here, we investigate V2-to-V1 coarticulation when there is no intervening boundary between the two vowels and when V1 is in IP absolute initial position (i.e. it is immediately preceded by an IP prosodic boundary, #V1C(C)V2 as in '*#essaie de faire ça*' [ɛsɛdfɛRsa] - 'try to do that', with # marking an IP boundary), a position known to show initial strengthening. Note that in this case, V1 is by definition in an onsetless syllable, that is, a position where we expect the vowel to be *prosodically strengthened* [11], thereby probably showing more resistance to contextual influence. Cases of V1 in IP-initial position are compared to a position where strengthening is not expected: medial position in both IP and Word (e.g. *ça épaissit* [saepɛsi] 'it thiken').

The second factor of interest in this study is the duration of the V1 target vowel. On the basis of pioneering work on *target undershoot* (cf. [12]), we know that many sources of contextual variation in continuous speech can influence the

http://dx.doi.org/10.21437/Interspeech.2016-1323

acoustic properties of a vowel. For mid-open and mid-closed vowels, factors favoring target undershoot would modify the F1 of V1 in the same direction than V2 coarticulatory would do in a harmonic context: lowering of F1 for mid-high vowels and raising of F1 for mid-closed vowels. In a large database of uncontrolled natural speech, such as the one we are using here, it is quite difficult to disentangle all the factors that could favor V1 undershoot (speech rate, style, prosody, lexical factors, among other ones). However, it is relatively easy to extract segment duration, a known co-variant of undershoot, whose effect on vowel acoustic properties has been relatively well described [13].

In the current study, we will thus indirectly test if the coarticulatory effect of V2 on V1 is influenced by other variation factors leading to V1 undershoot. Specifically, we look at whether V2-to-V1 coarticulation is stronger when V1 is temporally reduced.

2. Corpus-based study

Speech material was extracted from two publicly available French corpora: the corpus ESTER comprises continuous speech based on broadcasted news; the corpus NCCFr represents conversational speech from free and guided discussions on societal topics between friends [14-16].

2.1. Data coding

Words in which harmony could occur were selected according to the following principles: words containing one of the following vowels {i, e, ɛ, o, a, y, u, ɑ} as a V2 located in the last syllable and containing one of the following vowels {e, ɛ, o, ɔ} as a V1 placed in the penultimate syllable. A total of 33 325 words were selected by applying no constraint on their syllable structure, their duration and sequence of intervocalic consonants (i.e. between V1 and V2).

The vowels V1 and V2 were defined according to a forced automatic alignment, taking both orthographic and phonological information into account at the word level to raise possible ambiguities between mid-high and mid-close vowels. Vowels /ø, œ/ which associate to complex letter-to-phoneme relationships were excluded from the analysis. The mid V1 vowels were classified according to their place of articulation (/e, ɛ/ front vs. /o, ɔ/ back). These vowels were additionally coded according to whether the respective graphemes induce or not a higher pronunciation, that is, *é* for [e] and *au/eau* for [o].

The degree of V1-to-V2 coarticulation was measured as the lowering of the first formant of the V1 target in relation to the height of V2. Formant values were extracted at 1/3, 1/2 and 2/3 of the vowel and then averaged for a single value per vowel. In line with a previous study on vowel harmony [5], the statistical analysis was performed only on V1 represented by mid-front and mid-back vowels and on V2 including /+high/ (i.e. mid-high and high) vs. /-high/ (i.e. mid-low and low) vowels. The height of V2 is referred to as *V2 height*, a variable containing "high" and "low" values.

Other relevant variables comprise information related to the prosodic position of V1 (*Prosodic position*) and the duration of V1 (*V1 duration*). Concerning prosodic position, IP-initial position ("IPi") was determined on the basis of automatic pause detection: IP boundaries have been defined by the presence of pauses, and V1s immediately following a pause were labelled as "IPi". Therefore, all V1 in "IPi" are word initial (in an onsetless syllable) and are post-pausal. Since IP boundaries without pauses can also occur in French, it was not possible to confidently define the prosodic position of the other non-post-pausal word-initial V1. These cases were thus excluded and only Word-medial V1 (for which we are confident to say that they are IP-medial) were included in the analysis and coded as "Wm".

Duration was treated as a categorical variable made of two values values: "short" vowels (up to 50 ms) and "long" (above 50 ms) vowels, largely in line with previous classification criteria applied to the same corpora [13].

2.2. Statistical analyses

We used R [17] and lme4 package [18] to perform linear mixed effect analyses of the relationship between the first formant of V1 (in Hertz, F1 henceforth) and the following fixed effects:

- *V2 height*: "high" vs. "low" vowels
- *Prosodic position*: "IPi" vs. "Wm"
- *V1 Duration*: "long" vs. "short" vowels

Since style has been found to have a more subtle effect on V-to-V coarticulation than the one described in the literature [5], type of corpus (ESTER vs. NCCFr) was not treated as a factor in the analysis. The fixed structure of the model contained two interactions: one between *V2 height* and *Prosodic position* and one between *V2 height* and *V1 duration*. Given the very unbalanced frequency distribution of the relevant variables, we could not investigate further interactions between the various predictors. As random effects, we modelled intercepts for speakers. To avoid high Type I error rate [19], by-speaker random slopes (i.e. by-speaker variability on how fixed factors affect F1) were additionally fitted. Random-slopes were, however, modelled for the effect of V2 height only, due to a problem of convergence.

P-values were obtained on the basis of *Satterthwaite* approximations by using the *lmerTest* function [20], which gives a more conservative estimate of *p*-values for linear regressions. Main effects of each predictor and of interactions were tested by comparing the model with a certain factor (or interaction) with a model that lacks that particular factor (or interaction) by using Likelihood ratio test as implemented in the *anova*()-function. Finally, R^2 values associated to each model have been calculated by using the *r.squaredGLMM*()-function in [21], whose output provides R squared values. Specifically, this function calculates R^2 marginal values (R^2m) associated with fixed effects of the model and conditional R squared values (R^2c), which accounts for the proportion of variance explained by both fixed and random effects of the model.

3. Results

The linear mixed effect model (R^2m= .05 vs. R^2c= .33) shows a main effect of V2 height ($\chi^2(1)$=42.8, p<.0001), of Prosodic position ($\chi^2(2)$=92.7 p<.0001), and no effect of V1 duration (p=.17) on the F1 value of V1. As illustrated in Figure 1 and 2, V1 height (as measured by F1) is lowered or raised according to the height of V2 as predicted by V2-to-V1 coarticulation (or vowel height harmony) in French. More importantly, the model reveals:

- an interaction between V2 height and Prosodic position ($\chi^2(1)$=42.8, p<.0001): the effect of V2 height on the F1 of V1 is larger when V1 is in Wm position (β=35.1, SE=5.35, t=6.55, p<.0001, Figure 1). In other terms, V1 is more resistant to V2 coarticulation in IPi compared to Wm position;
- no interaction between V2 height and V1 duration (p=.08): In Figure 2 the effect of V2 is shown to be similar for both long and short V1.

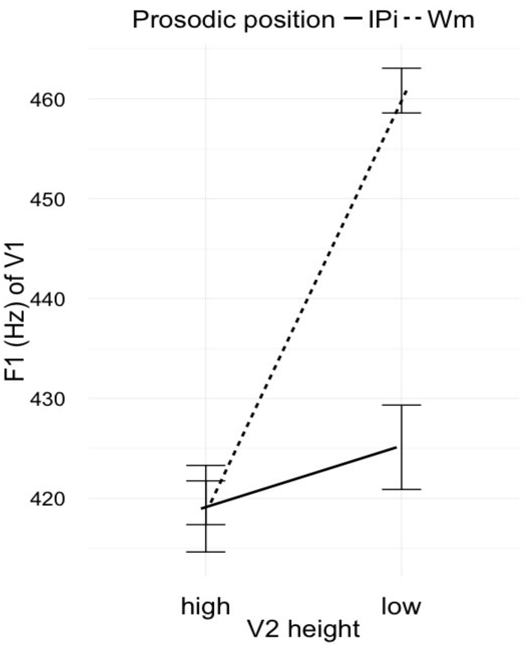

Figure 1: *F1 of V1 (as fitted by the model) as function of V2 height (high vs. low) and Prosodic position (IP initial vs. Word medial). Degree of vowel-to-vowel coarticulation can be inferred from the stiffness of the slope: greater in Wm. Whiskers represent standard errors.*

4. Discussion

The current study on French V-to-V coarticulation showed that the anticipatory effect from an upcoming vowel was smaller for target vowels sitting in a prosodically strong (IP absolute initial) position than for vowels located in a low-level (word medial) position.

These results go in line with the literature on coarticulatory variation across boundary (cf. Introduction). Prosodic-domain initial segments appear to be more resistant to coarticulation. To the extent that this coarticulatory resistance may be interpreted as being "edge"-induced or "boundary"-induced (cf. Introduction), the present results add to the previous literature in one important way. They suggest that initial positions in prosodically strong constituents are protected in a way that the prosodically strengthened segmental content is more resistant to contextual variation. Phonetic details linked to vowel distinctiveness and vocalic contrasts have been observed to be enhanced for vowels in prosodically strong position (e.g., [11], see also references therein for other languages than French) and have been taken as cues for prosodic strengthening. In light of the present results, a slightly reversed interpretation could be argued for: initial segments do not show an 'expansion' or a 'strengthening' of their spatio-temporal properties as often assumed in the literature, but rather a preservation of their distinctive, context-free, properties.

As it stands, the current study acts as a springboard for future work on prosodic effects of vowel-to-vowel coarticulation. While here we could focus our analysis on domain positions at the opposite ends of the hierarchical structure (IPi vs. Wm), in a follow-up study we are planning to conduct a more detailed acoustic and articulatory-based comparison between different prosodic positions on a controlled data-set.

Unlike prosodic boundary, the current study showed that vowel duration did not have any impact on V-to-V coarticulation: V2-to-V1 coarticulation was not stronger with temporally reduced V1. This finding goes along with previous work on effects of speaking rate ([22], [23]) and phonemic length contrast [24] on V-to-V (regressive) coarticulation. In these studies, none of these temporal effects was found to be correlated with coarticulatory variation. Furthermore, the absence of segment duration effects on V-to-V coarticulation suggests that this potential source of variation does not represent a confounding factor for the prosodically-induced coarticulatory variation that we observed here.

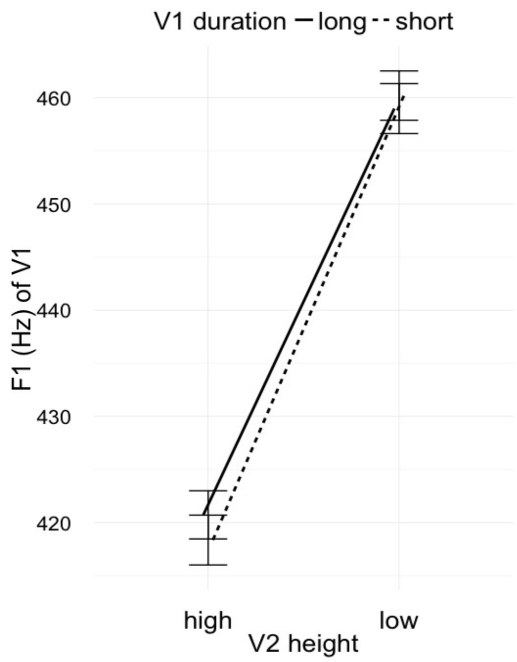

Figure 2: *F1 of V1 (as fitted by the model) as function of V2 height (high vs. low) and V1 duration (long vs. short). Degree of vowel-to-vowel coarticulation can be inferred from the stiffness of the slope: similar for long and short vowels. Whiskers represent standard errors.*

5. Conclusions and Outlook

This study investigated coarticulatory variation resulting from prosodic variation in French, a language where V-to-V coarticulation may result in an *optional* process of harmony. The goal was twofold: a) comparing degree of coarticulation (i.e. coarticulatory resistance) of vowels located at high vs. low-order prosodic level (IP absolute initial position vs. Word medial position); b) rule out the possibility that segment duration could be a confounding factor for vowel-to-vowel coarticulation. Though being widely acknowledged as a crucial factor of coarticulation, this study showed that duration does not exert the same influence on vowel-to-vowel coarticulation that prosodic position does.

Future articulatory-based investigations are expected to complement the current findings by looking at how prosodic boundary influences the detailed operation of the articulators involved in non-local coarticulatory interactions. This contribution would help to deepen our understanding on the still weak connection between articulatory accounts of motor control and psycholinguistic models of speech production.

Furthermore, future work should also look at the impact of these findings on speech comprehension. While we know that vowel harmony has a perceptual relevance for French listeners [25], we still do not know whether its interaction with prosodic structure would be based on the same mechanisms by which phonetic detail modulated by domain-initial strengthening assists listeners in lexical disambiguation (e.g., [26] for Dutch).

6. Acknowledgements

This research was funded by a research programme and innovation of the European Union Horizon 2020, through a Marie Sklodowska-Curie n° 662530 awarded to the first author and through the program “Investissements d’Avenir” overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL).

7. References

[1] F. Dell, *Les règles et les sons: Introduction à la phonologie générative.* (1972), Paris: Hermann.

[2] C. Fouché, *Traité de Prononciation Française*. 1956, Paris: Klincksieck.

[3] B. Tranel, *The Sounds of French: An Introduction*. 1987, Cambridge: CUP.

[4] N. Nguyen and Z. Fagyal, "Acoustic aspects of vowel harmony in French". *Journal of Phonetics*, 2008, 36(1), pp.1-27.

[5] G. Turco, C. Fougeron, and N. Audibert. "Que nous apprennent les gros corpus sur l’harmonie vocalique en français ?". To appear In *JEP 2016 - 31èmes Journées d’Etudes sur la Parole*. July 4-8, Paris, France, 2016.

[6] T. Cho, "The phonetics-prosody interface in laboratory phonology" In *Laboratory phonology*. The Continuum Companion to Phonology. , N.C. Kula, B. Botma, and K. Nasukawa, Editors. 2011, Continuum: London/New York:.

[7] C. Fougeron and P.A. Keating, "Articulatory strengthening at edges of prosodic domains". *The Journal of the Acoustical Society of America,* 1997, 101(6), pp. 3728-3740.

[8] E. Farnetani and D. Recasens, "Coarticulation models in recent speech production theories". In *Coarticulation: Theory, data and techniques*, W.J. Hardcastle and N. Hewlett, Editors. 1999, Cambridge University Press: Cambridge, UK. pp. 31-65.

[9] T. Cho, "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English". *Journal of Phonetics*, 2004, 32(2), pp. 141-176.

[10] B.G. Fivela, A. Stella, S. D'Apolito, F. Sigona. "Coarticulation Across Prosodic Domains in Italian: An Ultrasound Investigation". In *INTERSPEECH 2011 - 12th conference of the International Speech Communication Association*, August 27-31, Florence, Italy, 2011, pp. 393-396.

[11] L. Georgeton and C. Fougeron, "Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation". *Journal of Phonetics,* 2014, 44, pp. 83-95.

[12] B. Lindblom, "A spectrographic study of vowel reduction". *Journal of the Acoustical Society of America*, 1963, 37, pp. 1773-1781.

[13] N. Audibert, C. Fougeron, C. Gendrot, M. Adda-Decker. "Duration- vs. style-dependent vowel variation: A multiparametric investigation". In *ICPhS 2016 - 18th International Congress of Phonetic Sciences*, August 10-14, Glasgow, Scotland, 2016 (http://www.icphs2015.info/Proceedings.aspx).

[14] S. Galliano, G. Gravier, Chaubard, L. "The ESTER 2 Evaluation Campaign for the Rich Transcription of French Broadcast News". In *INTERSPEECH* 2009 - *10th conference of the International Speech Communication Association*, Brighton, UK, pp. 2583-2586.

[15] S. Galliano, E. Geoffrois, G. Gravier, J.-F. Bonastre, D. Mostefa, K. Choukri. "Corpus description of the Ester evaluation campaign for the rich transcription of French broadcast news". In *LREC (Vol.6) - 10th Language Resources and Evaluation Conference*, May 22-28, 2006, Genoa, Italy, pp. 315-320.

[16] F. Torreira, M. Adda-Decker, and M. Ernestus, "The Nijmegen corpus of casual French". *Speech Communication* 2010, 52, pp. 201-212.

[17] R Development Core Team, R: A language and environment for statistical computing. 2008, R Foundation for Statistical Computing: Vienna.

[18] D. Bates, et al., "lme4: Linear mixed-effects models using eigen and s4." R package version 1.1-7. http://CRAN.Rproject. org/package=lme4., 2014.

[19] I. Cunnings, "An overview of mixed-effects statistical models for second language researchers". *Second Language Research*, 2012, 28(3), pp. 369-382.

[20] A. Kuznetsova, P.B. Brockhoff, and R. H. B. Christensen. "Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)." R package version 2.0-29, 2013.

[21] S. Nakagawa and H. Schielzeth, "A general and simple method for obtaining R2 from generalized linear mixed-effects models". *Methods in Ecology and Evolution*, 2013, 4(2), pp. 133-142.

[22] M. Grosvald, "Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation". *Journal of Phonetics*, 2009, 37(2), pp. 173-188.

[23] I. Hertrich and H. Ackermann, "Coarticulation in slow speech: durational and spectral analysis". *Language and Speech*, 1995, 38(2), pp. 159-187.

[24] P.P. Mok, "Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation". *Language and Speech*, 2011, 54(4), pp. 527-545.

[25] N. Nguyen, Z. Fagyal, and J. Cole. "Perceptual relevance of long-domain phonetic dependencies". In JEL 2004 - *4èmes Journées d'Etudes Linguistiques (JEL)*, May 5-7, Nantes, France, 2004, pp. 173-178.

[26] T. Cho, J.M. McQueen, and A.E. Cox, "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English". *Journal of Phonetics*, 2007, 35(2), pp. 210-243.