



**HAL**  
open science

## Speech Style Effects on Local and Non-local Coarticulation in French

Giuseppina Turco, Fanny Guitard-Ivent, Cécile Fougeron

► **To cite this version:**

Giuseppina Turco, Fanny Guitard-Ivent, Cécile Fougeron. Speech Style Effects on Local and Non-local Coarticulation in French. Studies on Speech Production 11th International Seminar, ISSP 2017, Tianjin, China, October 16-19, 2017, Revised Selected Papers), pp.121-133, 2018, 10.1007/978-3-030-00126-1\_12 . hal-04337449

**HAL Id: hal-04337449**

**<https://hal.science/hal-04337449v1>**

Submitted on 12 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Please cite as : Turco, G., Guitard-Ivent, F., Fougeron, C. (2018). Speech Style Effects on Local and Non-local Coarticulation in French. In: Fang, Q., Dang, J., Perrier, P., Wei, J., Wang, L., Yan, N. (eds) *Studies on Speech Production*. ISSP 2017. Lecture Notes in Computer Science, vol 10733. Springer, Cham. [https://doi.org/10.1007/978-3-030-00126-1\\_12](https://doi.org/10.1007/978-3-030-00126-1_12)

## Speech style effects on local and non-local coarticulation in French

Giuseppina Turco, Fanny Guitard-Ivent and Cécile Fougeron

Lab. de Phonétique et Phonologie, UM7018, CNRS/Sorbonne Nouvelle

`giuseppina.turco-fanny.ivent-cecile.fougeron@univ-paris3.fr`

**Abstract.** Interactions between speech style and coarticulation are investigated by examining non-local, local, anticipatory and carryover contextual effects on vowels in two French corpora of conversational and journalistic speech. C-to-V coarticulation is analyzed on 22k tokens of /i, E, a, u, ɔ/ (/E/=e, ε/) falling within the same range of duration. Contextual effects are measured as F2 changes in relation to the adjacent consonant (alveolar vs. uvular) in CV<sub>1</sub> and V<sub>1</sub>C sequences. V-to-V coarticulation is analyzed on 33k V<sub>1</sub>C(C)V<sub>2</sub> sequences with V<sub>1</sub>=/e, ε, o, ɔ, a/ 50-to-80 ms long, and V<sub>2</sub> either high/mid-high or low/mid-low. Contextual effects are measured as F1 changes as function of V<sub>2</sub> height. Results show more local C-to-V coarticulation in conversational than in journalistic speech, as previously found for other languages. Interestingly, this interaction is clearer for all vowels in V<sub>1</sub>C, whereas coarticulation in CV<sub>1</sub> is affected by style for non-high vowels only. V-to-V coarticulation is also found in both corpora but is modulated by style only for mid-front vowels and in the opposite direction (i.e. more overlap in journalistic than in conversational speech). Findings are interpreted in light of dynamic models of speech production and of a phonological account of French V-to-V harmony.

**Keywords:** local vs. non-local coarticulation, speech style, French, large corpora, acoustic, anticipatory vs. carryover coarticulation

### 1 Introduction

Pronunciation is highly sensitive to speech style variations. In the literature, reported acoustic changes concern the spectral and temporal properties of a sound. Specifically, previous studies have shown that segment lengthening (Ferguson and Kewley-Port 2002, 2007; Picheny *et al.*, 1986; Moon and Lindblom 1994) and hyperarticulation, often measured with an expansion of the vowel acoustic space (Bradlow *et al.*, 2003; Johnson, Flemming and Wright 1993; Picheny *et al.*, 1986; Audibert *et al.*, 2014; Gendrot and Adda-Decker 2005), represent robust cues of “clear” (overarticulated) speech compared to other forms of elicited speech (citation-form speech) as well as casual speech (often referred to as spontaneous speech).

Less understood is the relation between speech style and coarticulation. Some studies revealed a reduction of coarticulation in clear speech (Krull 1989 for Swedish; Duez 1992 for French; Moon and Lindblom 1994 for English; DiCanio *et al.*, 2015 for Mixtec) compared to less clear speech (citation-form or casual-form). Other studies did not find any adjustments in degree of coarticulation according to speech style (Matthies *et al.*, 2001; Bradlow 2002). A third type of (more perception-oriented) study (Scarborough and Zellou 2013) found that it is the authenticity of the communicative context to induce variation in coarticulation degree (i.e. more coarticulation in a “real” than in a “simulated” clear speech situation).

The contradiction hiding behind all of these studies may be due to several reasons. First of all, they test different directions of coarticulation reflecting two distinct underlying processes: anticipatory coarticulation (as in Krull 1989; Duez 1992, Scarborough and Zellou 2013) vs. carryover coarticulation (as in Moon and Lindblom 1994; Bradlow 2002; DiCanio *et al.*, 2015). While both coarticulation can result from gestural overlap, anticipatory coarticulation can also reflect the planning of upcoming speech units, while carryover coarticulation can result from mechanical or inertial effects of moving an articulator from one target to the next one (e.g. Recasens, 1987).

Second, it is questionable whether the elicited speech conditions tested in these studies (from clear speech to more spontaneous speech) are comparable. Finally, there are cross-linguistic differences in coarticulatory patterns (Öhman 1966; Manuel 1990; Beddor 2002; Ma *et al.*, 2009) that may reflect differences in the relation between style and coarticulation across languages.

The current study investigates potential effects of speech style on vowel coarticulation in two big French corpora of (natural) casual conversation and of (natural) formal journalistic speech. Different from previous work mostly carried out on elicited speech, we investigate the relation between speech style and coarticulation in a more ecologically-valid speaking condition.

The novel aspect of the investigation lies in the comparison between CV<sub>1</sub> sequences with potential carryover coarticulation, V<sub>1</sub>C sequences with anticipatory coarticulation, and V<sub>1</sub>C(C)V<sub>2</sub> sequences with V-to-V anticipatory coarticulation. Hence, with respect to prior research mainly focusing on contextual overlap operating at a local level, here the observed coarticulation varies in terms of *distance* (local vs. non-local) and *cohesion* between the segments in context. The inter-articulator coordination within CV<sub>1</sub> sequences (coupled in-phase) has proven to be more stable than in (anti-phase) V<sub>1</sub>C sequences (Kelso *et al.*, 1986, Tuller and Kelso 1991, Nam *et al.*, 2010). Compared to C-to-V, V-to-V is the least cohesive structure, in which coordination seems to be the most language-dependent.

Another novel aspect of our study is the comparison among different sets of vowels. C-to-V coarticulation is analyzed on tokens of /i, E, a, u, ɔ/ (with /E/=e, ε/). These are the vowels for which we had enough material available in our two corpora in V<sub>1</sub>C and CV<sub>1</sub> sequences. Furthermore, the selection of this set of vowels allows us to test at the same time back vs. front vowels as well as high vs. non-high vowels. On the basis of the studies cited above, it is not possible to do predictions for each vowel. We expect the degree of coarticulation between C and V be modulated by speech style. At the same time, we set to determine whether the relation between coarticulation and style depends on the type of vowel. According to the Degree of Articulatory Constraints (DAC) model, vowels articulated with a palatal constriction are said to be more constrained articulatorily, and therefore, more resistant to contextual variation than others (Recasens 2007).

V-to-V coarticulation (in V<sub>1</sub>C(C)V<sub>2</sub> sequences) is tested on V<sub>1</sub> containing mid-front and mid-back vowels (≠e, ε, o, ɔ/) as well as the low vowel (=a/). While low vowels are particularly noteworthy for their high degree of contextual variation (e.g., Recasens 1987), mid-front and mid-back vowels are known to undergo a phonological process of anticipatory vowel harmony, at least for the Northern varieties of French (which are closer to Standard French, see Nguyen and Fagyal 2008; Turco, Fougeron and Audibert, 2016). The target V<sub>1</sub> usually alternates in height (mid-high/mid-low) according to the height changes of the following V<sub>2</sub> (high vs. low, respectively). The classic literature on French vowel harmony (e.g., Fouché 1959, Walker 2001) suggests that degree of influence of V<sub>2</sub> on V<sub>1</sub> may be sensitive to speech style variations. V<sub>1</sub> harmonizes more with V<sub>2</sub> in casual than in formal speech, an observation that is largely consistent with previous findings on local coarticulation and speech style. However, these observations are still far from being systematically tested (see Turco *et al.*, 2016 for a preliminary investigation). In the current study, we aim at investigating the relation between vowel harmony and style more systematically.

A further open question in the literature is whether variations affecting the spatial (formant) properties of the vowel are induced by variations in duration. In a revised version of Lindblom (1963)'s vowel *undershoot* model, Moon and Lindblom (1994) showed that vowel formant displacements due to consonantal context highly depended on vowel duration and that the link between spectral changes and duration was less strong in clear-speech than in citation-form speech (a similar finding is also reported by Duez 1992 for V-to-C coarticulation in French). However, we know that speech style variation is more than just a matter of duration: speakers can adopt different production strategies according to the speaking condition at hand (e.g. Ferguson and Kewley-Port, 2002; Smiljanic and Bradlow, 2005). As a matter of fact, previous work on speech style differences in French has shown that vowel spectral reduction in conversational speech compared to read speech mainly applied when looking at short vowels (under 40ms; see, for instance, Rouas *et al.*, 2010). Moreover, Audibert *et al.* (2015) found more reduction of the acoustic space, more vowel centralization and intra-category dispersion in conversational speech than in journalistic and read speech for short vowels (up to 50ms) than for mid (up to 80ms) and long vowels (up to 120ms). Hence, on the basis of these findings, we test potential effects of speech style on coarticulation by controlling for the duration of the target vowel.

To summarize, the questions we address in the current study are the following ones:

1. whether degree of coarticulation varies with speech style (and, if so, in which direction and for which vowels);
2. whether speech style variation equally affects all types of coarticulation (anticipatory vs. carryover, local vs. non-local).
3. whether speech style variation affects vowel coarticulation when its duration is controlled for.

## 2 A corpus-based study

Speech material was extracted from two publicly available French corpora: ESTER (Gravier *et al.*, 2006) and NCCFr (Torreira *et al.*, 2010). ESTER contains broadcasted news and political/societal debates of several radio and TV programs. It includes mainly scripted speech mostly produced by professional speakers. NCCFr contains conversational speech based on free and guided face-to-face discussions on societal topics among pair of friends (mainly young students). In our study, the corpus ESTER will be referred to as “journalistic” speech (J, henceforth), the corpus NCCFr as “conversational” speech (C).

### 2.1 Speech material and Data coding

Local C-to-V coarticulation was observed on 22k tokens extracted from both corpora. The target vowel ( $V_1$ ) was represented by the following set of sounds: /i, E, a, u, ə/ (with /E/=e, ε/), produced by 23 male speakers. The duration of the target vowel was controlled: all vowels ranged from 50 to 80 ms. The adjacent consonant was either an alveolar ( $C_{ALV}$ =/t, d, z, s, l, n/, e.g. *dépanner* /depane/ - “to help”), which is known to attract F2 towards an 1800Hz locus, or a uvular ( $C_{UV}$ =/R/, e.g. *appareil* /apaRej/ - “appliance”), known to lower F2 (and raise F1). The opposite context (left in  $V_1C$  and right in  $CV_1$ ) was always a labial consonant. Note that in  $CV_1$  sequence, C and  $V_1$  are always tautosyllabic, while in  $V_1C$  sequences syllabification could not be controlled for. Contextual effects were measured as changes of the second formant (F2) of  $V_1$  according to the place of articulation of the adjacent C.

V-to-V coarticulation was analyzed on 33k words containing a  $V_1C(C)V_2$  sequence where the target  $V_1$  was a mid-high, mid-low or low vowel (/e, ε, o, ə, a/) located in the penultimate syllable of the word. The influencing vowel  $V_2$  was either a raising trigger (high and mid-high /i, e, o, y, u/, all coded as ‘high’) or a lowering trigger (low and mid-low /ε, a, ā/, all coded as ‘low’) placed in the last (accented) syllable. In the extraction of those words, no constraint on their syllable structure and their sequence of intervocalic consonants (i.e. between  $V_1$  and  $V_2$ ) was applied. The degree of coarticulation was measured as the lowering of the first formant (F1) of  $V_1$  in relation to the height of  $V_2$  (high/mid-high vs. low/mid-low).

The vowels were identified according to a forced automatic alignment (using the Burg algorithm in the Praat software, Boersma & Weenink, 2018) that takes into account both orthographic and phonological information at a word level (see Turco *et al.* 2016 for details). In all cases, the duration of  $V_1$  was controlled to avoid duration-dependent variation in coarticulation degree. All vowels were between 50ms and 80ms. These duration values were obtained following classification criteria applied to the same corpora as described in Audibert *et al.* (2015). Finally, 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 order to check for outliers (due, for instance, to mislabelling), formant values were inspected according to the procedure described in Gendrot and Adda-Dekker (2005). They were then *z*-score transformed (Lobanov 1971) by subtracting the mean of all data points from each of them and by dividing those points by the standard deviation of all data points.

### 2.2 Statistical analyses

We performed linear mixed-effects models using the lme4 package (Bates *et al.*, 2015) in the R software.

Two types of model structure were constructed. For local C-to-V coarticulation, we modeled the relationship between the F2 of the target  $V_1$  (in *z*-score) in relation to the adjacent CONSONANTAL CONTEXT (C: alveolar vs. uvular), DIRECTION TYPE ( $CV_1$  vs.  $V_1C$ ) and SPEECH STYLE (conversational vs. journalistic speech). For non-local V-to-V coarticulation, the model contained the F1 of  $V_1$  (in *z*-score) as function of  $V_2$  HEIGHT (high vs. low) and SPEECH STYLE (conversational vs. journalistic speech). In both models, speaker was included as a random factor. By-speaker random slopes were also included to avoid high Type I error rate (Cunnings, 2012). The two models were then run separately for each vowel according to the type of coarticulation they belonged to (C-to-V or V-to-V). *P*-value estimates were based on *Satterthwaite* approximations, which provides more conservative estimates for linear regression, through the *lmerTest*()-function (Kuznetsova *et al.*, 2013) and further adjusted for multiple testing comparisons. Likelihood ratio tests as implemented in the *anova*()-function were performed to check main effects of each fixed factor and interactions.

In line with previous studies (cf. Introduction), we predicted more coarticulation in conversational than in formal journalistic speech. In statistical terms, this is translated into an interaction between the relevant predictors (i.e. between CONSONANTAL CONTEXT and SPEECH STYLE for  $CV_1$  and  $V_1C$  sequences;  $V_2$  HEIGHT and SPEECH STYLE for  $V_1CCV_2$  sequences). In what follows, we present only the relevant effects and interactions that reply to our research questions.

### 3 Results

#### 3.1 C-to-V coarticulation

First, the statistical analyses reveal a significant main effect of CONSONANTAL CONTEXT on the F2 of all the tested vowels (see Table 1 below). As expected, the F2 of the target vowels is higher when these vowels are in alveolar than in a uvular context.

**Table 1.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the effect of the predictor CONSONANTAL CONTEXT on the F2 of five vowels (/i/, /u/, E=/e,  $\epsilon$ /, /ɔ/, /a/). The intercept contains “alveolar” as a reference value; “n.s.” stands for  $p$ -values that are not significant.

CONSONANTAL CONTEXT				
	$\beta_{\text{uvular}}$	SE	$t$	$p$ -value
/i/	-.245	.055	-4.42	.0006
/u/	-.390	.056	-6.91	.0001
/E/	-.090	.031	-2.88	.01
/ɔ/	-.380	.073	-5.17	.0001
/a/	-.471	.048	-9.90	.0001

By contrast, unexpectedly, an effect of DIRECTION TYPE is found for the vowels /E/, /ɔ/ and /a/: the F2 of these vowels is higher in CV<sub>1</sub> sequences than in V<sub>1</sub>C sequences (see Table 2).

**Table 2.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the effect of the predictor DIRECTION TYPE on the F2 of five vowels (/i/, /u/, E=/e,  $\epsilon$ /, /ɔ/, /a/). The intercept contains “VC” sequence as a reference value; “n.s.” stands for  $p$ -values that are not significant.

DIRECTION TYPE				
	$\beta_{\text{cv}}$	SE	$t$	$p$ -value
/i/	.017	.016	1.08	<i>n.s.</i>
/u/	.089	.061	1.44	<i>n.s.</i>
/E/	.254	.027	9.25	.001
/ɔ/	.263	.042	6.27	.0001
/a/	.148	.033	4.50	.0001

Furthermore, our analyses reveal an effect of SPEECH STYLE on F2 for peripheral vowels /i, a/ (Table 3): their F2 is higher in journalistic speech than in conversational speech.

**Table 3.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the effect of the predictor STYLE on the F2 of five vowels (/i/, /u/, E=/e,  $\epsilon$ /, /ɔ/, /a/). The intercept contains “journalistic” as a reference value; “n.s.” stands for  $p$ -values that are not significant.

STYLE				
	$\beta_{\text{conversational}}$	SE	$t$	$p$ -value
/i/	.136	.048	2.83	.009
/u/	-.096	.089	-1.08	<i>n.s.</i>
/E/	.085	.034	2.48	.02
/ɔ/	.250	.086	2.90	.008
/a/	.008	.063	.13	<i>n.s.</i>

More interestingly, an interaction between CONSONANTAL CONTEXT and SPEECH STYLE is found for all the tested vowels: there is less coarticulation in journalistic speech than in conversational speech (Table 4).

**Table 4.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the interaction between CONSONANTAL CONTEXT\*SPEECH STYLE on the F2 of five vowels (/i/, /u/, E=/e,  $\epsilon$ /, /ɔ/, /a/). The intercept contains alveolar and journalistic as reference values; “n.s.” stands for  $p$ -values that are not significant.

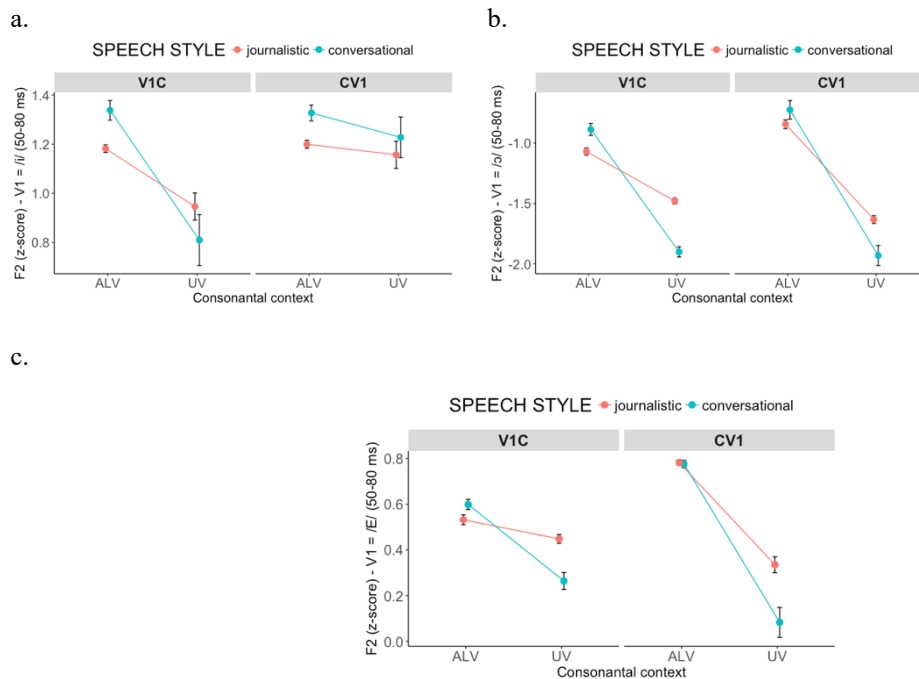
CONSONANTAL CONTEXT*SPEECH STYLE				
	$\beta_{\text{uvular*conversational}}$	SE	$t$	$p$ -value
/i/	-.265	.080	-3.31	.003

/u/	-.359	.077	-4.62	.0001
/E/	-.238	.041	-5.75	.0001
/ɔ/	-.714	.093	-7.67	.0001
/a/	-.344	.059	-5.80	.0001

Finally, the models reveal an interaction between CONSONANTAL CONTEXT, SPEECH STYLE and DIRECTION TYPE for the vowels /i, u, a, ɔ/ (see Table 5). Indeed, for close vowels /i, u/, the interaction is found in anticipatory V<sub>1</sub>C coarticulation only (see Figure 1a). For non-close vowels /ɔ, a/, this effect is present in both V<sub>1</sub>C and CV<sub>1</sub> sequences. Yet, the reduction of coarticulation in journalistic speech is more important in V<sub>1</sub>C than in CV<sub>1</sub> sequence (see Figure 1b). In other words, SPEECH STYLE seems to affect mostly anticipatory coarticulation. For /E/, DIRECTION TYPE does not affect the interaction between CONSONANTAL CONTEXT and SPEECH STYLE: less coarticulation in journalistic speech is found in both V<sub>1</sub>C and CV<sub>1</sub> sequence, as illustrated in Figure 1c.

**Table 5.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the interaction between CONSONANTAL CONTEXT\*DIRECTION TYPE\*STYLE on the F2 of five vowels (/i/, /u/, E=/e,  $\epsilon$ /, /ɔ/, /a/). The intercept contains alveolar, VC and journalistic as reference values; “n.s.” stands for  $p$ -values that are not significant.

CONSONANTAL CONTEXT*DIRECTION TYPE*STYLE				
	B <sub>uvular*CV*conversational</sub>	SE	$t$	$p$ -value
/i/	.216	.090	2.40	.03
/u/	.328	.143	2.30	.03
/E/	-.020	.065	-.30	n.s.
/ɔ/	.277	.111	2.49	.02
/a/	.187	.065	2.88	.009



**Fig. 1.** F2 (in z-score) of the vowels /i/ (panel a), /ɔ/ (panel b) and /E/ (panel c) ranging from 50 to 80 ms in V<sub>1</sub>C and CV<sub>1</sub> context as function of consonantal context (alveolar vs. uvular) and speech style (journalistic vs. conversational).

### 3.2 V-to-V coarticulation

For V<sub>1</sub>C(C)V<sub>2</sub> sequences, the models reveal that V-to-V coarticulation is found for all the tested V<sub>1</sub> in the expected direction: F1 is lower when followed by a high V<sub>2</sub> (see Table 6 for model output). However, the change in F1 conditioned by V<sub>2</sub> HEIGHT is larger in magnitude for the mid-back vowels /o, ɔ/ and for the low /a/ than for the mid-front vowels /e,  $\epsilon$ /.

**Table 6.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the effect of the predictor V<sub>2</sub> HEIGHT on the F1 of three vowels (/a/, E=/e,  $\epsilon$ /, O=/o,  $\text{ɔ}$ /). The intercept contains “high” as a reference value; “n.s.” stands for  $p$ -values that are not significant.

V <sub>2</sub> HEIGHT				
	$\beta_{\text{low}}$	SE	$t$	$p$ -value
/a/	.069	.004	17.50	.0001
/E/	.032	.004	7.60	.0001
/O/	.124	.006	20.41	.0001

The main effect of STYLE is found for the vowel /a/ only, showing a higher F1 in journalistic speech than in conversational speech (cf. Table 7). Together with the higher F2 for /i/ and the lower F2 for /u/ described in the previous section (cf. 3.1), the higher F1 of /a/ in journalistic speech contributes to the expansion of the vowel acoustic space expected in journalistic speech.

**Table 7.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the effect of the predictor STYLE on the F1 of three vowels (/a/, E=/e,  $\epsilon$ /, O=/o,  $\text{ɔ}$ /). The intercept contains “journalistic” as a reference value; “n.s.” stands for  $p$ -values that are not significant.

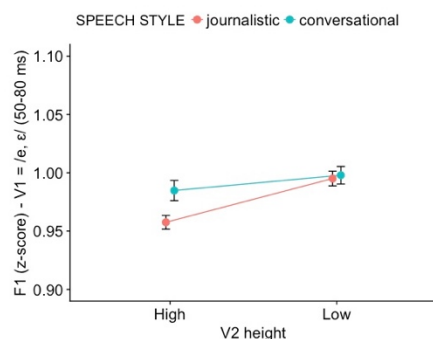
STYLE				
	$\beta_{\text{conversational}}$	SE	$t$	$p$ -value
/a/	-.041	.014	-2.90	.006
/E/	-.009	.014	-.63	n.s.
/O/	-.009	.016	-.56	n.s.

Crucially, the models reveal an interaction between V<sub>2</sub>HEIGHT and STYLE for the E=/e,  $\epsilon$ / only (cf. Table 8). For the mid-front V<sub>1</sub>, there is less coarticulation in conversational speech than in journalistic speech (see Figure 2a). On the contrary, there is much more contextual change of F1 for the low and mid-back vowels, but this holds true for both speech styles (Figure 2b and 2c), as indicated by the lack of a statistical interaction.

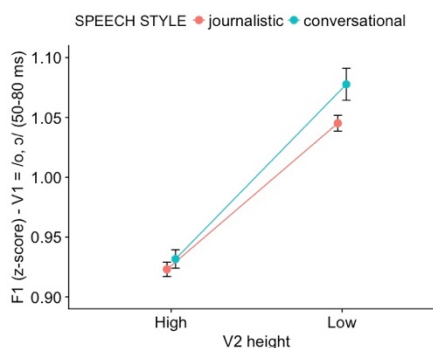
**Table 8.** Estimates ( $\beta$ -coefficients, standard errors (SE),  $t$ -values and  $p$ -values) of the linear mixed effects models for the interaction between the predictor V<sub>2</sub> HEIGHT\*STYLE on the F1 of three vowels (/a/, E=/e,  $\epsilon$ /, O=/o,  $\text{ɔ}$ /). The intercept contains “high” and “journalistic” as reference values; “n.s.” stands for  $p$ -values that are not significant.

V <sub>2</sub> HEIGHT*STYLE				
	$\beta_{\text{low*conversational}}$	SE	$t$	$p$ -value
/a/	.008	.006	1.29	n.s.
/E/	-.017	.006	-2.66	.01
/O/	.003	.011	.29	n.s.

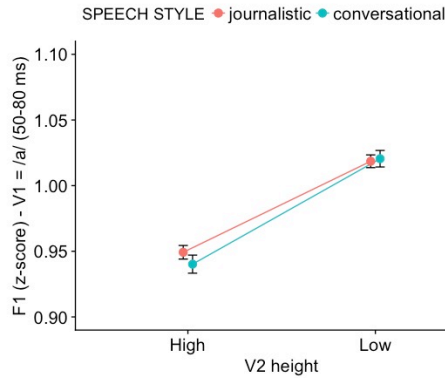
a.



b.



c.



**Fig. 2.** F1 (in z-score) of the vowels ( $V_1$ ) /e,  $\epsilon$ / (panel a), /o,  $\text{o}$ / (panel b) and /a/ ranging from 50 to 80 ms as function of the height (high vs. low) of the following vowel ( $V_2$ ) and speech style (journalistic vs. conversational).

## 4 Discussion and conclusion

Based on large French corpora of journalistic and conversational speech, in the current study we set to determine 1) whether the degree of segment overlap varies with speech style (and, if so, in which direction); 2) whether local carry-over  $CV_1$ , anticipatory  $V_1C$  and non-local anticipatory  $V_1C(C)V_2$  vary with style (i.e. whether the effect of style on the magnitude of coarticulation in the three cases tested here can tell us something about the difference between these three processes); 3) whether vowel coarticulation is affected by style when its duration is controlled for.

In line with previous studies (Krull 1989; Duez 1992; Moon and Lindblom 1994; DiCanio *et al.*, 2015), for C-to-V coarticulation it was found that degree of overlap changed according to speech style – vowels were less coarticulated in formal (journalistic) than in informal (conversational) speech; for V-to-V, the opposite pattern was observed – there was some more coarticulation in formal speech than in informal speech. This interaction occurred for the mid-front vowels only, that is, for that set of vowels undergoing a phonological process of vowel harmony. For mid-back and low /a/ vowel, coarticulation occurred regardless of style differences (and was even stronger compared to mid-front vowels).

The question is why are there differences in the direction of style effect between local and non-local coarticulation? While our findings are consistent with preliminary findings carried out on the same set of data (Turco, Fougeron and Audibert 2016), they do not confirm previous observations found in the classic literature for French (cf. Introduction). One explanation for why mid-front vowels assimilate less in informal than formal speech may be simply because we are dealing with a phonological phenomenon: vowel height changes affecting mid-front vowels in French may be considered as the reflex of a phonological process, and as such it shows crucial differences from V-to-V coarticulation. This phonological account finds also indirect evidence in the study by Nguyen and Fagyal (2008). The authors noticed that, during the laboratory recording session, their Southern French speakers changed their (Southern French) colloquial style into a more (Northern-like French) formal reading style. This “style-shift” led those participants to produce more vowel harmony in the tested words. It is furthermore interesting that this interaction occurs specifically for mid-front vowels (and not for mid-back vowels). Considering that our conversational corpus is made of productions by young speakers (unlike the journalistic corpus), it is possible that we are dealing with a phenomenon of diachronic change. Height distinctions of mid-front vowels are more neutralized in the speech of young than old Parisians. Note, however, that the effect of the interaction (despite differences in the direction) was not as strong as it turns out to be for the C-to-V coarticulation, which leads us to reply to our second question.

Our analyses revealed that the three types of coarticulation were not affected by style to the same extent. For local coarticulation, ( $V_1C$ ) anticipatory coarticulation varied more with style than carryover ( $CV_1$ ) coarticulation. Specifically, compared to  $C_1V$ ,  $V_1C$  showed less coarticulation in formal (journalistic) speech than in informal (conversational) speech for the set of vowels /i, u, O, a/, whereas for /E/ the degree of overlap modulated by style was the same across the two types of sequences. Moreover, compared to local coarticulation, there was less degree of overlap varying with style in non-local V-to-V coarticulation ( $V_1CCV_2$ ). So, the crucial question here is why  $CV_1$  or  $V_1C$  should vary – though differently – with style whereas V-to-V does not.

For local coarticulation, the ‘coupled oscillator’ model seems to best account for the differences observed between  $V_1C$  and  $CV_1$  sequences. Being characterized by a stable inter-articulator timing coordination, it is conceivable that a  $CV_1$  sequence (coupled in-phase) shows less sensitivity to speech style variation. By contrast,



based on an anti-phase coupling, in a V<sub>1</sub>C sequence the coordination of segments is less stable and hence may be more subject to potential sources of variation such as style (Kelso, Saltzman and Tuller, 1986; Tuller and Kelso; 1991, Nam, Goldstein and Saltzman, 2010). On the other hand, non-local V-to-V coarticulation (in extent, direction and magnitude) is language-dependent (Manuel 1990, 1999), speaker-dependent (Magen 1987) and, more importantly, dependent on the segment in the sequence over which is planned (Whalen, 1990). Variability in V-to-V coarticulation has been shown to be mainly due to the articulatory demands of the intervening consonant (Recasens 1982; Ohman 1966), that is, to mechanical restrictions on tongue movements (e.g. degree of tongue dorsum contact). Crucially, these mechanical restrictions have been found to be more important for V-to-V anticipatory coarticulation than for carryover coarticulation (Recasens, 1982) and seem to not vary with style. In Turco *et al.*, (2016), more transconsonantal coarticulation was indeed found with a less constraining consonant (i.e. labials, as also found by Recasens 1982 on V-to-V in Catalan) between the two mid-vowels (i.e. vowel harmony). By contrast, in the current study based on the same corpora as Turco *et al.*, (2016), a much larger variety of pivot consonants were included between the two vowels. It is hence likely that V-to-V coarticulation may remain unaffected by style variation since the planning mechanism (i.e. anticipating the second trigger vowel during the first target vowel) on which V-to-V is based, depends essentially on the nature of the intervening consonant, and not on style or other factors like speech rate and stress (Recasens, 2015).

Finally, our findings suggest that the effect of speech style on coarticulation is not a by-product of durational variation. Added to the previous body of research on effects of style on the acoustics of vowels with controlled duration (Rouas *et al.*, 2013 and Audibert *et al.*, 2015); we have been able to show effects of style on degree of coarticulation of vowels falling within a certain range of duration (50-80ms). In light of these findings, it therefore seems important to consider style as a factor of variation in its own right. Future studies should test if similar or different effects apply to other categories of vowel duration and if so, their implications for speech production models.

## 5 Acknowledgements

This research was funded by a research program and innovation of the European Union Horizon 2020, through a Marie Skłodowska-Curie n° 661530 awarded to the first author and by the ANR-10-LABX-0083 (Labex EFL).

## 6 References

1. Audibert, N., Fougeron, C., Gendrot, C., and M. Adda-Decker (2015). "Duration- vs. style-dependent variation: a multiparametric investigation." *Proceedings of the 2015 ICPhS conference*. Glasgow (<https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0753.pdf>)
2. Bates, D. Maechler, M. Bolker, B and S. Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67 (1), pp. 1-48 ([doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)).
3. Beddor, P. S., Harnsberger, J. D., and S. Lindemann (2002). "Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates." *Journal of Phonetics*, 30 (4), pp. 591-627.
4. Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 11 May 2018 from <http://www.praat.org/>
5. Bradlow, A. R. (2002). "Confluent talker-and listener-oriented forces in clear speech production." *Laboratory phonology*, 7, pp. 241-272.
6. Cunnings, I. (2012). "An overview of mixed-effects statistical models for second language researchers." *Second Language Research*, 28(3), pp. 369-382.
7. Dicanio, C., Nam, H., Amith, J. D., García, R. C., and D.H. Whalen (2015). "Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec." *Journal of Phonetics* 48, pp. 45-59.
8. Duez, D. (1992). "Second formant locus-nucleus patterns: An investigation of spontaneous French speech." *Speech Communication*, 11(4-5), pp. 417-427.
9. Ferguson, S. H., and D. Kewley-Port (2002). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners." *The Journal of the Acoustical Society of America*, 112(1), pp. 259-271.
10. Fouché, P. (1959). *Traité de prononciation française*. Paris: Klincksieck.
11. Gendrot, C., and Adda, M. (2005) Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. In *Proceedings of Eurospeech*. Lisbon, Portugal. 2453–2456.
12. Gravier, G., Bonastre, J-F., Geoffrois, E., Galliano, S., Tait, K. Mc., and K. Choukri (2006). "Corpus description of the ESTER evaluation campaign for the rich transcription of French broadcast news." *Proceedings of European Conference on Speech Communication and Technology*, pp. 139–142.

13. Kelso, J. A., Saltzman, E. L., and B. Tuller (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14 (1), pp. 29-59.
14. Krull D. (1989). "Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech." *Perilus* 10, pp. 87-108.
15. Kuznetsova, A., Brockhoff, P.B., and R. H. B. Christensen (2013). "LmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)." R package version 2. pp 0-29.
16. Lobanov, B. M. (1971) Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49, 606-608.
17. Ma, L., Perrier, P., and J. Dang (2006). "Anticipatory Coarticulation in Vowel-Consonant-Vowel sequences: A crosslinguistic study of French and Mandarin speakers." *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba. pp. 151-158.
18. Manuel, S. Y. (1990). "The role of contrast in limiting vowel-to-vowel coarticulation in different languages." *The Journal of the Acoustical Society of America*, 88 (3), pp. 1286-1298.
19. Matthies, M., Perrier, P., Perkell, J. S. and M. Zandipour (2001). "Variation in anticipatory coarticulation with changes in clarity and rate." *Journal of Speech Language and Hearing Research* 44, pp. 340-353.
20. Meunier, C., and R. Espesser (2012). "Vowel reduction in conversational speech in French: The role of lexical factors." *Journal of Phonetics* 39 (3), pp. 271-278.
21. Moon, S. J., and B. Lindblom (1994). "Interaction between duration, context, and speaking style in English stressed vowels." *The Journal of the Acoustical Society of America*, 96 (1), pp. 40-55.
22. Nam, H., Goldstein, L. and Saltzman, E. (2010). "Self-organization of syllable structure: A coupled oscillator model." In Pellegrino, F. (Ed.) *Approaches to phonological complexity* (Vol. 16). Walter de Gruyter, Berlin, pp. 299-328.
23. Öhman, S. E. (1966). "Coarticulation in VCV utterances: Spectrographic measurements." *The Journal of the Acoustical Society of America*, 39 (1), pp. 151-168.
24. R Development Core Team: A language and environment for statistical computing. 2008, R Foundation for Statistical Computing: Vienna ([www.r-project.org](http://www.r-project.org)).
25. Recasens, D. (1987). "An acoustic analysis of V-to-C and V-to-V: Coarticulatory effects in Catalan and Spanish VCV sequences." *Journal of Phonetics*, 15(1), 71-86.
26. Recasens, D. (2007). "Patterns of VCV coarticulatory direction according to the DAC model." In Prieto, P., Mascaró, J. & Solé, M.J. (eds.) *Segmental and Prosodic Issues in Romance Phonology*, John Benjamins, Amsterdam, pp. 25-40.
27. Rouas, J.L., Beppu, M., and M. Adda-Decker (2010). "Comparison of Spectral Properties of Read, Prepared and Casual Speech in French." *International Conference on Language Resources and Evaluation (LREC 2010)*, Malta, pp. 606-611.
28. Scarborough, R., and G. Zellou (2013). "Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception." *The Journal of the Acoustical Society of America*, 134(5), pp. 3793-3807.
29. Smiljanić, R., and A.R Bradlow (2005). "Production and perception of clear speech in Croatian and English a". *The Journal of the Acoustical Society of America*, 118(3), pp. 1677-1688.
30. Torreira, F., Adda-Decker, M., and M. Ernestus (2010). "The Nijmegen corpus of casual French." *Speech Communication* 52, pp. 201-212.
31. Tranel, B. (1987). *The Sounds of French: An Introduction*. Cambridge: CUP.
32. Tuller, B., and J. A. S. Kelso (1991). "The production and perception of syllable structure." *Journal of Speech and Hearing Research*, 34, pp. 501-508.
33. Turco, G., Fougeron, C., and A. Audibert (2016). "Que nous apprennent les gros corpus sur l'harmonie vocalique en français ?" *Proceedings of the 2016 JEP-TALN-RECITAL*, Paris, pp. 571-579.
34. Walker, D. C. (2001). *French Sound Structure*. Calgary: University of Calgary Press.