STRESS CONDITIONED PHONOLOGICAL PROCESS: A CASE STUDY OF ITALIAN PALATALIZATION

[1,2] Bowei Shao, [2] Philipp Buech, [2] Anne Hermes, [1] Maria Giavazzi

[1] Département d'études cognitives, École Normale Supérieure, Université PSL, Paris, France
[2] Laboratoire de Phonétique et Phonologie, UMR7018, CNRS/Université Sorbonne Nouvelle, Paris, France
{bowei.shao; maria.giavazzi}@ens.psl.eu, {philipp.buech; anne.hermes}@sorbonne-nouvelle.fr

ABSTRACT

This study examines the interaction between lexical prominence and palatal/velar consonants in Italian. The specific phenomenon of interest is the palatalization of /k/ in, e.g., 'cardiaci' [kar.ˈdi.a.**tʃ**i] but not in 'ubriachi' [u.bri.ˈa.**k**i], which is argued to be conditioned by the position of /k/ relative to lexical prominence. The aim of the study is to uncover potential mechanisms that block palatalization in post-stress position, as in [u.bri.ˈa.**k**i]. The results revealed that the stressed vowel /a/ is more peripheral, i.e., hyper-articulated compared to unstressed /a/, which is more closed and fronted, especially towards its end. Further, the immediately post-stress consonants show longer closure durations compared to far-from-stress. We argue that the blocking of velar palatalization could be related to the longer closure duration in post-stress context, which is an articulatory by-product of the hyper-articulated stressed vowel. These findings are discussed within the predictions of the Task Dynamic model for the *μ-gesture.*

Keywords: Lexical prominence, stress, Italian palatalization, *μ*-gesture, acoustics.

1. INTRODUCTION

Prosodically prominent positions within the word are known (1) to be privileged, manifesting the positional maintenance of contrasts otherwise neutralized and the resistance to processes applying elsewhere. For instance, vocalic contrasts are preferentially realized in stressed position, but they may be reduced in unstressed position (e.g., in English, Brazilian Portuguese, Western Catalan [1, 2, 3]). Similar patterns are observed for consonantal contrasts (e.g., in Copala Trique, Italian, and Finnish [4, 5, 6]). Prominent positions are however also (2) the preferred target for a small class of frequent processes: for instance, consonants are often lengthened in pre-tonic and post-tonic positions (as in English and Somali [7, 8]), consonants in these two positions are described as having louder (e.g., Farsi [9, 10]) or affricate-like bursts (e.g., Maori [11]).

The two seemingly contradictory linguistic behaviours, (1) and (2), may have a common phonetic basis: they may be the articulatory by-product of lexical prominence. The nucleus bearing lexical prominence is known to be longer in duration, higher in f0, more peripheral and more intense [12]. While f0, duration and intensity are less related to oral cavity activities, the hyper-articulated vowel qualities related to lexical prominence shall manifest on the lingual articulation of the nucleus and have effects on adjacent segments [13, 14]. This interaction between lexical prominence and segmental processes can be illustrated by velar palatalization in Italian masculine plural nouns and adjectives. This process is triggered by /i/, and conditioned by lexical prominence [15, 16, 4]. For example:

(3) FAR: cardiaci /kar.ˈdi.a.k-i/→[kar.ˈdi.a.tʃi] 'cardiac'
POST: ubriachi /u.bri.ˈa.k-i/→[u.bri.ˈa.ki] 'drunk'

In (3) FAR, stress is located on the second syllable of the word and considered here as far from the target of palatalization /k/ (hereafter FAR), the /k/ is palatalized to /tʃ/. However, in (3) POST, palatalization is blocked in the immediate post-stress position (hereafter POST). Overall, the examples in (3) illustrate the triggering and the blocking of palatalization in Italian. The triggering of palatalization by the following /i/ is argued to be both acoustically and auditorily motivated [17, 18, 19]: the velar stop /k/ and the palato-alveolar affricate /tʃ/ before a palatal vowel /i/ share acoustic similarities, which have been analyzed as the cause of perceptual reanalysis by speakers/listeners. The reason for the blocking of palatalization, however, is still unclear.

In this study, we aim to investigate how the blocking of palatalization could be explained in relation to lexical prominence by analyzing the acoustic characteristics of stressed vowels and the following segments in two stress conditions (POST and FAR). The hypothesis is that a tempo-spatial modulation gesture related to lexical prominence, i.e., *μ*-gesture [14], may have played an important role in modulating the articulatory behaviour of both the stressed vowel and the consonant in POST condition.

This prominence related gesture modulation, in turn, has an impact on the perceptual contrast between /k, g/ *vs.* /tʃ, dʒ/.

2. METHOD

2.1. Material and speakers

Target words were trisyllabic nonce words, structured /$C_1V_1.C_2V_2.C_3V_3$/, differing solely by the position of stress on the first or the second syllable (e.g., /ˈpi.ta.ki/, /pi.ˈta.ki/). The nonce words were designed to compare how the target consonants /k, g, tʃ, dʒ/ (in C_3 position) were produced in both FAR and POST contexts. The V_3 position is occupied by /i/. $C_1V_1.C_2V_2$ sequences were /pita/, /fesa/, /pufa/ /tipa/ and /suta/. For example, in the nonce words /ˈpi.ta.ki/ and /pi.ˈta.ki/, C_3 /k/ was in FAR and POST contexts respectively.

The nonce words were written in their orthographic forms with stress location marked by an acute lexical stress accent (e.g., *pítachi, pitáchi*). They were embedded in a carrier phrase *Dimmi* ___ *di nuovo* ('Say ___ again'). The phrases were randomized and presented on a computer screen.

The acoustic recordings were conducted in a double-walled sound booth on a TASCAM DR-100MKIII Linear PCM Recorder. Sound files were recorded at a sampling rate of 44.1 kHz. The phrases were repeated three times by 18 speakers, yielding 3328 tokens (two speakers were excluded due to misplacement of stress and dialectal background).

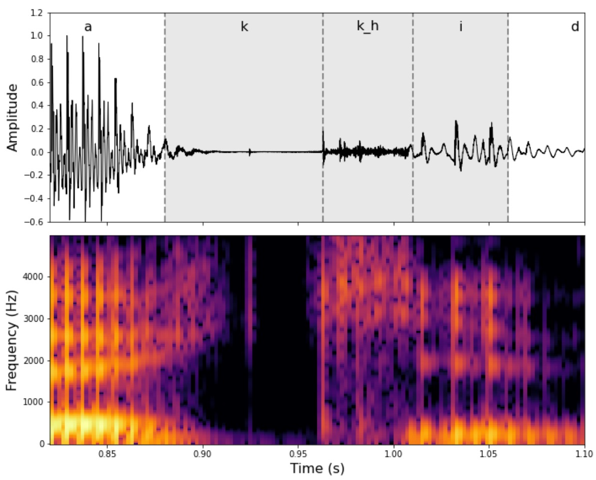

Figure 1: Segmentation of the target sequence '*áchi*' in the utterance '*Dimmi pitáchi di nuovo'*.

Utterances were pre-segmented using the Label Sounds function of Audacity v3.1.3 [20]. They were then extracted and automatically annotated using the Montreal Forced Aligner v2.0.6 [21]. Stop bursts were automatically identified by custom script written in Python v.3.10.6 [22] utilising a modified version of the algorithm described in [23]. The alignments (see Fig. 1) were manually inspected and corrected.

Acoustic analyses were conducted on V_2, C_3 and V_3 using Praat (6.1.16) [24]. We measured the duration and the formants for V_2 and V_3. Formants were measured by taking ten equal-distanced time points across the entire duration of the vowels (including onset and offset, with limits set to 5000 Hz for male and 5500 Hz for female speakers), in order to account for the dynamic evolution during the vowels. For C_3, we measured the mean Centre of Gravity (CoG) using windows of 5 ms shifted by 1 ms across the release phase (i.e., the aspiration phase of /k, g/ and the fricative phase of /tʃ, dʒ/). The closure duration and total duration of the plosives and affricates was also measured for C_3. We calculated the closure ratio by taking the ratio between closure duration and total duration of C_3.

2.2. Statistical analyses

The formant trajectories and the duration of V_2 were modelled with generalized additive mixed models (GAMM) [25] using the *mgcv* package (1.8-40) [26] in *Rstudio* [27]. The model was constructed as shown in the footnote[1]. With this model, we were able to analyze how formant trajectories vary as a function of stress, and how the time-varying formant trajectories were affected by the duration of the vowels.

To compare the differences in CoG and in closure duration across C_3, we conducted linear mixed-effects analyses, using *Rstudio* [27] and the *lme4* (1.1-30) [28] package. Mean CoG/closure ratio were taken as dependent variables, and the interaction between C_3 and stress as fixed effects. As random effects, we had by-speaker random slopes for the effect of C_3. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

3. RESULTS

3.1. Duration and formant structure of V_2

The stressed vowel is known to be hyper-articulated, entailing *inter alia* longer duration [12]. The durational difference between stressed and unstressed V_2 is evident, with the stressed V_2 being more than twice as long ($\bar{x}$ = 185 ms, σ = 45) as the unstressed V_2 ($\bar{x}$ = 82 ms, σ = 29).

The vocalic spaces of F1 and F2 (in Hz) of V_2 are shown in Fig. 2. The stressed vowel is more open, the unstressed vowel is less open, fronted and clearly more variable.

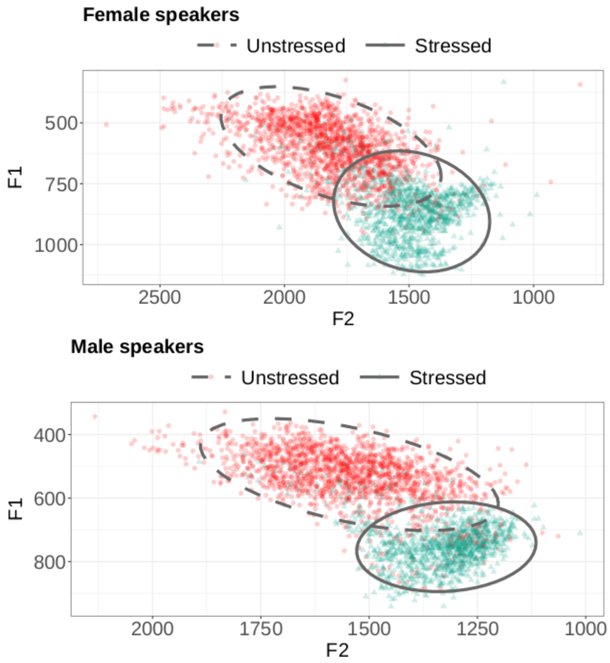

Figure 2: Formants (F1 and F2 in HZ) of /a/ according to stress conditions (red/dashed lines for unstressed; green/solid for stressed), with 95% confidence ellipses; female speakers on top, male speakers below.

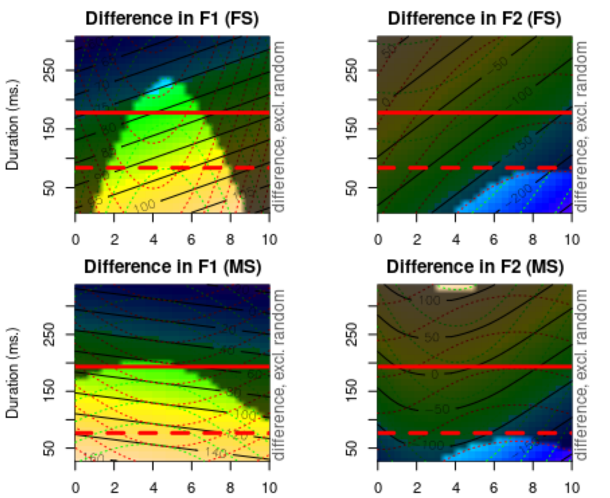

Figure 3: Difference between stressed /a/ and unstressed /a/ over time (x-axis) modulated by duration (y-axis). The shaded region indicates the area where the difference is non-significant. The red solid and dashed lines indicate the mean duration of stressed and unstressed /a/ respectively. FS and MS are for female and male speakers respectively.

The duration and formant structure of V_2 is further explored with GAMMs. More precisely, the models investigate how the formant trajectories of V_2 evolve in relation to the difference in duration. That is, does a longer, stressed V_2 differ from a shorter, unstressed V_2 in terms of formant trajectory? This comparison is shown in Fig. 3, which presents the estimated differences between stressed and unstressed V_2. The general observation would be that the F1 is overall higher in stressed /a/, except at the beginning and the end, probably due to coarticulation with surrounding segments. The models estimate that the difference is larger when the duration is shorter. This could be understood as follows: the shorter the unstressed vowel /a/, the more closed it is. The difference in F2 is less pronounced compared to F1. However, there is a consistent pattern towards the end of the vowels: the models predict that the shorter the vowel, the higher its F2.

In general, the differences between stressed and unstressed V_2 show a duration-related pattern. When V_2 is short, it is less open over its entire duration, and particularly fronted towards its end.

3.2. Closure duration and mean centre of gravity in C_3

The mean CoG of a consonant indicates its place of articulation in the oral cavity [29]. The CoG in C_3 is analysed with linear mixed effect models, grouped by their voicing categories. The results are presented in Table 1.

Centre of gravity	Estimate	S.E.	*p*
/k/ POST (Intercept)	3896.1	126.2	***
/tʃ/ POST	-88.1	120.1	0.476
/k/ FAR	-6.7	33.2	0.841
/tʃ/ FAR	-29.7	47.1	0.528
/g/ POST (Intercept)	1442.7	111.0	***
/dʒ/ POST	1350.1	137.0	***
/g/ FAR	-124.7	85.1	0.144
/dʒ/ FAR	103.2	121.4	0.396
Closure ratio			
/k/ POST (Intercept)	0.574	0.017	***
/g/ POST	0.080	0.035	*
/k/ FAR	-0.044	0.012	***
/g/ FAR	0.055	0.017	**
/tʃ/ POST (Intercept)	0.371	0.015	***
/dʒ/ POST	0.117	0.013	***
/tʃ/ FAR	-0.028	0.009	**
/dʒ/ FAR	0.030	0.013	*

Table 1: Fixed effects of linear mixed-effects analyses conducted on CoG and closure ratio of C_3, grouped by voicing and manner of articulation, respectively. *=<0.5, **<0.01, ***<0.001.

We note that CoG of plosives and affricates are overall similar, especially for the voiceless ones. The release phase of /k/, whether in POST or in FAR, has similar CoG compared to /tʃ/ in both contexts. The /g/ in POST has lower CoG compared to /dʒ/ in POST, but in FAR both segments are not significantly different from each other and from /g/ in POST. These results show that CoG is not related to the relative position of lexical stress; it is most probably related to the coarticulation with the following vowel /i/, as shown in [17].

The results of two mixed models on closure ratio, grouped by manners of articulation, are presented in Table 1. The important general pattern is that, for all consonants, the closure phase is more important in POST than in FAR. This difference is more pronounced in the plosives compared to the affricates.

3.3. Formant structure and duration of V_3

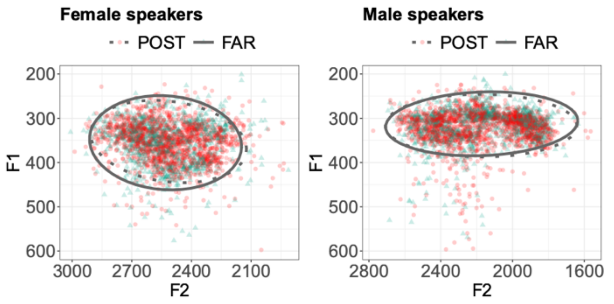

Figure 4: Formant (F1 and F2 in Hz) of /i/ (V_3) according to position (red/dotted line for POST, green/solid line for FAR), with 95% confidence ellipses.

V_3 /i/ is analysed in relation to its relative position to lexical stress. The formant structures of /i/ show no difference, as can be seen in Fig. 4. The F1/F2 spaces for /i/ in POST and FAR contexts are virtually identical. The duration of V_3 showed no difference either (POST: $\bar{x} = 73$ ms, $\sigma = 42$; FAR: $\bar{x} = 86$ ms, $\sigma = 40$). This suggests that V_3 is not directly influenced by the position of lexical stress. Whether in POST or FAR, the acoustic characteristics of /i/ are virtually the same. This further suggests that /i/ has the potential of triggering palatalization in both conditions, however palatalization is not realised in POST.

4. DISCUSSION AND CONCLUSION

This study investigated the influence of lexical prominence on the following segments in a /$C_1V_1.C_2V_2.C_3V_3$/ sequence in Italian, aiming to uncover potential mechanisms that block velar palatalization in post-stress position. By putting stress on V_1 or on V_2, the syllable C_3V_3 was either FAR from stress or POST stress. The analysis on stressed vowels showed that the μ-conditioned stressed vowels were longer in duration and more peripheral in acoustic space. The analyses of C_3V_3 in both conditions showed that V_3 (i.e., /i/ and trigger of palatalization) could not be responsible for the blocking of velar palatalization as it had similar acoustic characteristics in both FAR and POST conditions. However, C_3 had longer closure duration in POST compared to in FAR. This difference, combined with the acoustic shape of μ-conditioned V_2, may shed light on the stress-conditioning of palatalization.

We interpret the stress-conditioned velar palatalization as resulting from a perceptual effect rooted in articulation. The durational increase of the stressed vowel follows the predictions of the μ-gesture model [14]. As illustrated in Fig. 5, given that gestures are temporally coordinated and potentially overlapping with other gestures, μ-gestures also have an extent in time and overlap with vocal tract constriction gestures, namely here the C_3 closure gesture. The intergestural coarticulatory overlap between two adjacent movements, i.e., an opening gesture of V_2 and a closure gesture of C_3, are realized in different degrees according to the presence or absence of the μ-gesture (i.e., POST and FAR conditions) [30].

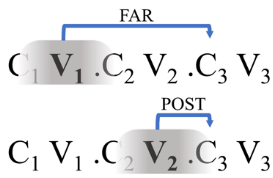

Figure 5: Illustration of μ-gesture associated with different stressed positions. Shading represents the scope of the μ-gesture, with strongest effect at the midpoint of the stressed vowel.

In POST, the μ-conditioned V_2 shows more coarticulatory resistance [31] with the following C_3. The large opening gesture of V_2 leads to a later target achievement of C_3, reflected acoustically in the longer closure duration. In FAR, the μ-gesture is not having an effect. The V_2 gesture shows less coarticulatory resistance, resulting in more gestural blending and a shorter closure in C_3.

The articulatory-conditioned C_3 closure duration discussed above is likely to have played an important perceptual role in recovering the plosive *vs.* affricate category. That is, a longer closure duration may facilitate the recovering of a plosive consonant, leading to the blocking of palatalization.

In sum, the difference in C_3 closure duration is interpreted as an articulatory by-product directly related to the μ-conditioned V_2. To confirm our interpretation, an articulatory study is planned to fully understand the gestural coordination between μ-conditioned vowels and consonants.

5. ACKNOWLEDGEMENTS

We'd like to thank Sam Mitchell for his contribution in the study's earlier stage. This study was funded by ANR-JCJC DIA-SYN-PHON funding attributed to Maria Giavazzi (ANR-21-CE28-0008). This study was partially supported by EUR Frontiers in Cognition and the Idex PSL (ANR-17-EURE-0017), and Labex EFL (ANR-10-LABX-0083).

5. REFERENCES

[1] Fourakis, M. 1991. Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical society of America*, 90(4), 1816-1827.
[2] Marcet, A., Fernández-López, M., Baciero, A., Sesé, A., & Perea, M. 2022. What are the letters e and é in a language with vowel reduction? The case of Catalan. *Applied Psycholinguistics*, 43(1), 193-210.
[3] Nobre, M. A., & Ingemann, F. 1987. Oral vowel reduction in Brazilian Portuguese. *In Honour of Ilse Lehiste*, 195-206.
[4] Giavazzi, M. 2010. *The Phonetics of Metrical Prominence and its Consequences on Segmental Phonology*. Doctoral dissertation, MIT.
[5] Hollenbach, B. 1977. Phonetic vs. phonemic correspondence in two Trique dialects. In WilliamMerrifield (ed.), Studies in Otomanguean Phonolog. Dallas: SIL and University of Texas atArlington. 3567.
[6] Suomi, K., Ylitalo, R. 2004. On Durational Correlates of Word Stress in Finnish. Journal ofPhonetic, 32:35-63.
[7] Armstrong, L. E. 1964. *The phonetic structure of Somali*. Gregg International Publishers Limited.
[8] Turk, A. E., & Sawusch, J. R. 1997. The domain of accentual lengthening in American English. *Journal of Phonetics*, *25*(1), 25-41.
[9] Gonzalèz, C. 2003. *The effect of stress and foot structure on consonantal processes*. Doctoral dissertation, USC
[10] Samareh, Y. 1977. *Arrangement of segmental phonemes in Farsi*. University of Tehran, Tehran University Publications.
[11] Bauer, W. 1993. *Maori*. London, New York: Routledge.
[12] Gordon, M., & Roettger, T. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, 3(1), 1-11.
[13] Byrd, D., & Saltzman, E. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2), 149-180.
[14] Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. 2008. A Task-dynamic Toolkit for Modeling the Effects of Prosodic Structure on Articulation. In *Proceedings of the 4th International Conference on Speech Prosody (speech prosody 2008)*, Campinas, Brazil. 175-184.
[15] Dressler, W. U. 1985. *Morphophonology: The Dynamics of Derivation*. Ann Arbor, Mich.: Karoma.
[16] Albano Leoni, F., Cutugno, F., Savy, R., 1995. The vowel system of Italian connected speech, in Elenius K., Branderud P., ed., *Proceedings of the XIIIth ICPhS*, Stockholm, 4, 396-399
[17] Guion, S. G. 1998. The role of perception in the sound change of velar palatalization. *Phonetica*, 55(1-2), 18-52.
[18] Ohala, J. J., & Jones, C. 1993. The phonetics of sound change. *Historical linguistics: Problems and perspectives*, 237-278.
[19] Wilson, C. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science*, 30(5), 945-982.
[20] Audacity Team. 2022. *Audacity(R): Free Audio Editor and Recorder* [Computer application]. Version 3.1.3 retrieved December 25th 2021 from https://audacityteam.org/.
[21] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498-502.
[22] Van Rossum, G., & Drake, F. L. 2009. *Python/C Api Manual-Python 3*. CreateSpace.
[23] Ananthapadmanabha, T. V., Prathosh, A. P., & Ramakrishnan, A. G. 2014. Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index. *The Journal of the Acoustical Society of America*, 135(1), 460-471.
[24] Boersma, P. & Weenink, D. 2022. *Praat: doing phonetics by computer* [Computer program]. Version 6.1.16, retrieved 21 November 2020 from http://www.praat.org/.
[25] Wood, S. 2006. *Generalized additive models: an introduction with R*. CRC Press.
[26] Wood S. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3-36.
[27] R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[28] Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.
[29] Stevens, K. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.
[30] Mücke, D., & Grice, M. 2014. The effect of focus marking on supralaryngeal articulation–Is it mediated by accentuation?. *Journal of Phonetics*, 44, 47-61.
[31] Browman, C. P., & Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. Papers in laboratory phonology I: Between the grammar and physics of speech, 341-376.

[1] GAMM models

```
bam(valuef1 ~ stressPattern*duration_ms +
   s(pt_f1, by = stressPattern, k = 10) +
   s(pt_f1, speaker, by = stressPattern,
      bs = "fs", m = 1, k = 10) +
   s(pt_f1, cons, by = stressPattern,
      bs = "fs", k = 10) +
   s(pt_f1, cons_aft, by = stressPattern,
      bs = "fs", k = 10),
   data = fmt_melt_a_m,
   scat(theta = NULL, link = "identity",
       min.df = 3),
   nthreads = 30,
   discrete = T,
   AR.start = fmt_melt_a_m$start.event,
rho = fmt_autocorr)
```