Testing the structure preservation and phonetic preservation approaches to compensatory lengthening

Enkeleida Kapia[1,2], Josiane Riverin-Coutlée[1], Conceição Cunha[1], and Jonathan Harrington[1]

[1]*Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München (Germany),*
[2]*Albanian Academy of Sciences (Albania)*

This study tests predictions from two approaches to compensatory lengthening (CL) against empirical data from the Tosk dialect of Albanian. CL is a diachronic process defined as the lengthening of a target segment upon deletion of a trigger segment, for instance $C_1V_1C_2V_2$ > $C_1V_1{:}C_2$ [1]. Over the past 50 years, there has been a documented tendency in Tosk for $C_1V_1C_2V_2$ words like *kalë* 'horse' /ˈkalə/ to lose their final schwa, raising the question of whether CL could develop in this dialect [2]. Two approaches to CL allow to make predictions in this regard. The structure preservation approach (A1) suggests that CL will not develop in Tosk because it does not feature contrastive vowel length [3]. The phonetic preservation approach (A2), formulated by [4] for Slavic languages, conceptualizes CL as a set of mathematical operations applied to the physical duration of target and trigger segments. Once a target segment reaches 1.5 time or more its original duration, the threshold (arbitrarily) established by [4] for phonological reanalysis, CL is expected to develop.

In this study, we tested the predictions from A1 and A2 using a corpus wherein Tosk speakers variably produced conservative and innovative forms of $C_1V_1C_2V_2$ words, e.g. [ˈkalə] and [kal] respectively. Verifying the prediction from A1 was straightforward: if confirmed, V_1 would have similar durations in conservative and innovative words. To test A2, the mathematical equation in (1), adapted from [4], was applied to vowel duration in conservative words ([ˈkalə]) to verify if it could predict vowel duration in innovative words ([kal]) and whether this predicted duration reached the threshold where CL is expected to develop:

(1) $d(V_1 \text{ in } C_1V_1C_2) = d(V_1 \text{ in } C_1V_1C_2V_2) + \alpha - \beta - \gamma$

where d(x) is the duration of segment x; α is $d(V_2 \text{ in } C_1V_1C_2V_2)$; β and γ are penalizing values for the shortening effect of certain types of C_2 and V_1 heights respectively (more details below).

Sixteen speakers of Tosk were recorded during a reading task in which they produced four repetitions of 250 words, including 30 words traditionally ending with schwa (V_2). The forced-aligned [5] speech signals were hand-corrected, after which EMU-SDMS [6] was used to measure stressed V_1 duration, and duration of V_2 in conservative words (22% of the corpus, not lexically conditioned). Median V_2 duration (88 ms) was set as value for the α parameter in formula (1). To obtain values for the β and γ parameters, we extracted estimates from a linear mixed effect regression [7,8] modelling the effect of C_2 (sonorant, voiceless obstruent) and V_1 height (low, mid, high) on V_1 duration in conservative words, with random intercepts per speaker and word. A similar model integrating an additional two-level innovativeness factor tested for differences in V_1 duration between innovative and conservative words (A1).

The violins in Figure 1 represent measured V_1 durations across vowel heights and C_2 types in conservative words in gray, and innovative words in orange. Diamonds correspond to empirical means. Predictions for innovative V_1 obtained from formula (1) are represented by black circles and bars. In all cases, predictions (black) are much higher than real values (orange); they suggest CL will develop everywhere, with V_1 predicted to be 1.54 to 1.75 time longer in innovative than conservative words. In reality, however, V_1 duration in innovative and conservative words hardly differs (*est.*=3.11, $t(177)=1.68$, $p=0.09$), suggesting no emergent CL in Tosk. Our results are thus compatible with the prediction of A1 only, which has also found extensive support elsewhere [9]. This recurring tendency for CL to develop in languages with pre-existing length contrasts suggests an important role of phonological representations for its emergence [3,10], but also a different implementation mechanism than an exchange of physical duration between trigger and target, e.g. [11].

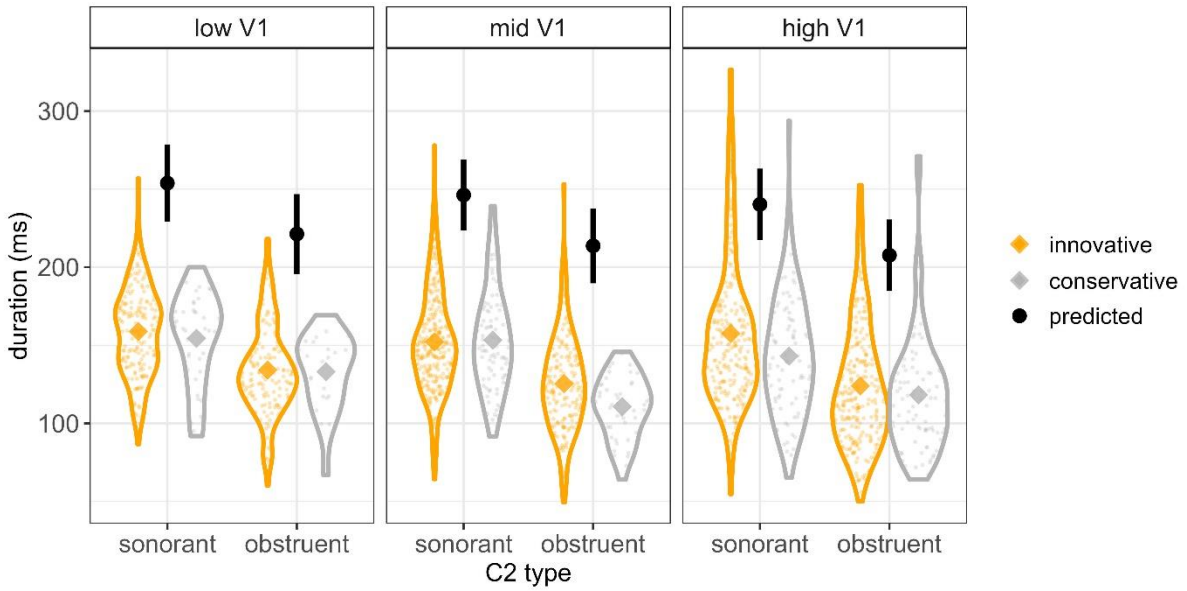

Fig. 1. Stressed vowel duration (ms) measured in innovative (orange) and conservative (gray) words, against predicted vowel duration in innovative words (black) obtained from formula (1), for two types of C_2 and three V_1 heights.

References

[1] Gess, R. (2011). Compensatory lengthening. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell Companion to Phonology*, vol. III (pp. 1536–1557). Oxford: Wiley Blackwell.

[2] Çeliku, M. (1971). Kuantiteti i zanoreve të theksuara në të folmet e shqipes. *Studime Filologjike*, 8(4), 65–100.

[3] de Chene, B. E. & S. R. Anderson. (1979). Compensatory lengthening. *Language*, 55(3), 505–535.

[4] Timberlake, A. (1983). Compensatory lengthening in Slavic, 2: Phonetic reconstruction. In M. S. Flier (ed.), *American Contributions to the 9th International Congress of Slavists*, vol. I (pp. 293–320). Bloomington: Slavica Publishers.

[5] Kisler, T., U. D. Reichel & F. Schiel. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347.

[6] Winkelmann, R., J. Harrington & K. Jänsch. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, 45, 392–410.

[7] Bates, D., M. Maechler, B. Bolker & S. Walker. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

[8] R Core Team. (2023). *R: A language and environment for statistical computing*. https://www.R-project.org.

[9] Kavitskaya, D. (2017). Compensatory lengthening and structure preservation revisited yet again. In C. Bowern, L. Horn & R. Zanuttini (eds.), *On Looking into Words (and Beyond)* (pp. 41–58). Berlin: Language Science Press.

[10] Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2), 253–306.

[11] Kavitskaya, D. (2001). *Compensatory Lengthening: Phonetics, Phonology, Diachrony*. University of California, Berkeley PhD dissertation.