INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Increasing aspiration of word-medial fortis plosives in Swiss Standard German

Franka Zebe[1]

[1]University of Zurich

franka.zebe@uzh.ch

Abstract

There is evidence for a sound change in progress in German-speaking Switzerland: Namely, Swiss German speakers of Alemannic increasingly use aspiration in fortis plosives, particularly in word-initial position. This study aims to extend the research by investigating word-medial plosives in Swiss Standard German (SSG). Using the apparent-time paradigm, the main goal is to compare younger to older speakers.

Since the increasing aspiration is probably driven by the contact to German Standard German (GSG), this study focuses on speakers from both rural and urban areas, assuming that the latter have more contact to speakers of GSG than the former.

Results show that younger urban speakers produce longer VOT values in alveolar plosives than the other speakers, while all younger speakers show this pattern for bilabial plosives. Furthermore, only the younger speakers from the urban group produce shorter closure durations in fortis plosives.

Index Terms: sound change, Swiss Standard German, aspirated plosives, VOT, closure duration

1. Introduction

Recent evidence suggests the existence of a sound change in progress in Alemannic dialects spoken in Switzerland [1 – 3]. Specifically, it has been shown that younger speakers produce fortis plosives in word-initial position with an increasing amount of aspiration [3], while both younger and older speakers use closure duration as primary cue in distinguishing between lenis (shorter closure) and fortis (longer closure) plosives. The use of aspirated fortis plosives is specifically distinctive in word-initial plosives, while in word-medial plosives, the age-related difference is much smaller [3]. At this point, the available evidence points to an externally driven sound change, induced by contact to German Standard German (GSG). Therefore, the current study aims to investigate, for the first time, both VOT and closure duration of word-medial plosives in Swiss Standard German (SSG) using the apparent-time paradigm [4, 5].

1.1. Swiss Standard German and Alemannic

German-speaking Switzerland is a typical example for a diglossia [6], where a low variety, Alemannic, and a high variety, SSG, are spoken by the community. In general, Alemannic is spoken in most every-day situations, while SSG is learned in school and spoken in formal situations. According to [7], the pronunciation of SSG is not necessarily uniform, depending, among other things, on the speaker's dialect [7, 8]. Since Alemannic is the main form of spoken communication and SSG is also influenced by the speakers' dialectal varieties, it is assumed that the diglossia in German-speaking Switzerland is relatively stable.

Yet, the majority of studies on Swiss German focus solely on Alemannic, including recent investigations on the ongoing sound change in German-speaking Switzerland [1, 3]. The aim of this study is to build on the research on Alemannic by investigating speakers' productions of word-medial fortis plosives in SSG, using the apparent-time paradigm [4, 5].

1.2. VOT and closure duration

It is evident that the main difference between lenis and fortis plosive consonants in Alemannic is closure duration, which is significantly longer for fortis compared to lenis plosives [9 – 12]. While speakers of SSG produce longer VOT durations in fortis compared to lenis plosives [7], closure duration remains the primary correlate of the contrast between homorganic stops [8, 13]. In contrast, GSG speakers use VOT as primary cue to differentiate between word-medial lenis and fortis plosives, while closure duration is a secondary cue [14].

Recent evidence shows that younger speakers of Zurich German produce shorter closure durations in word-medial fortis plosives in both Alemannic and SSG compared to older speakers [13]. At the same time, a recently conducted apparent-time analysis of Swiss speakers of Alemannic shows that fortis plosives are increasingly pronounced as aspirated by young speakers, particularly in word-initial position [3]. This trend is also found in word-medial position but it is of a much weaker magnitude [3]. Taken together, these results suggest the possibility of an increasing importance of VOT and a decreasing importance of closure duration. Since these two aspects have not been investigated together yet, this study focuses on both VOT and closure duration.

1.3. Sound change and language contact

Although the diglossic situation in German-speaking Switzerland is, in general, assumed to be stable, recent evidence shows that a sound change in progress is taking place in Alemannic [1 – 3]. This ongoing sound change is likely originated in the contact to GSG for two reasons. First, in GSG VOT is the primary cue to distinguish between lenis and fortis plosives, while closure duration is a secondary cue [14]. Furthermore, there is also evidence for an increase of the importance of VOT in other varieties in Austria and Germany, likely also due to the contact to GSG [15, 16]. Second, lower frequency words seem to be more affected than higher frequency words [3], which is in accordance with both the Frequency Actuation Hypothesis [17]. In addition, results add to the increasing amount of evidence for phonetically gradual sound changes due to language contact [18 – 20].

The current study aims to further investigate the possibility that the sound change is externally driven by focusing on both an urban area with higher contact to GSG and a rural area with lower contact to GSG. If the sound change is indeed externally driven, SSG can be argued to be the step in between GSG and Alemannic. It is expected to find a generational difference in

 10.21437/Interspeech.2023-295

the production of fortis plosives in SSG which should be more pronounced in the urban compared to the rural group of speakers. In addition, because plosives in word-medial position are much less affected by this sound change in Alemannic compared to word-initial ones [3], the current study focuses on the question whether the age-related difference in word-medial plosives is of a larger extent in SSG compared to Alemannic.

Ultimately, this study aims to answer the following research questions: (1) Do younger speakers use VOT more than older speakers to differentiate between word-medial lenis and fortis plosives in SSG? (2) Is this effect stronger for the urban compared to the rural group of speakers? (3) Is an increase of aspiration accompanied with a decrease in closure duration?

2. Methods

2.1. Speakers and stimuli

Of the 40 speakers recorded, 20 were from rural areas in the canton of Lucerne; at the time of the recording, 10 (5 female) of them were younger with ages between 25 and 32 years (mean=28.7, SD=1.94) and 10 (5 female) of them were older with ages between 47 and 64 (mean=58.0, SD=6.30). The other 20 speakers were from urban areas near or within the city of Zurich; 10 (5 female) belong to the younger group with ages between 18 and 28 (mean=23.0, SD=2.98), whereas 10 (5 female) of them belong to the older group with ages between 58 and 69 (mean=64.1, SD=3.75).

Stimuli consisted of 16 disyllabic words, 7 with a word-medial lenis plosive and 9 with a word-medial fortis plosive. The targets comprised the two lenis plosives /d̥/ and /b̥/ as well as the two fortis plosives /t/ and /p/. Since the plosive /k/ is often realized as the affricate /k͡x/ by Swiss German speakers [7], velar plosives are not included in this study. Each target plosive was preceded by the vowels /a/, /i/, or /u/ (short or long). All words were embedded in a carrier sentence consisting of seven syllables in total (e.g., Er will den *Kater* füttern. 'He wants to feed the *tomcat*.'). The primary stress of each sentence was on the first syllable of the target word.

2.2. Procedure and data preparation

Whenever possible, recordings were conducted in a soundproof booth at the Phonetics Laboratory of the University of Zurich using a personal computer with a USBPre® 2 (Sound Devices) interface and a NT2-A (RØDE) microphone. In some cases, participants were recorded in a different, quiet room using a laptop computer with the same interface and an Opus 54.16/3 (beyerdynamic) microphone.

To present the stimuli to the speakers, the SpeechRecorder software [21] was used. The participants were instructed to first read the sentences silently and then say them loud in their natural speech tempo. The sentences were presented in a semi-random order, so that the same stimuli never appeared twice in a row. The order was different for each participant. Each sentence was repeated five times, resulting in a total number of 3200 recordings (40 speakers x 16 words x 5 repetitions).

All recordings were automatically segmented using webMAUS [22]. Next, the automatic segmentations were manually corrected using the EMU-webApp [23]. The start and end segment of each sentence and each target word was adjusted. In addition, each segment of the target word was adjusted, while the word-medial plosives were separated into two segments, i.e., closure duration and release (VOT). The onset of the closure duration was defined as the end of a clearly visible second formant of the preceding vowel. The VOT segment started at the point of release of the closure duration and ended at a clearly visible second formant of the following vowel. To rule out possible effects of articulation rate, the absolute closure and VOT values were then normalized by dividing their duration by the absolute word duration, using the same procedure as in [18].

Some recordings had to be excluded from the analysis due to technical issues, long pauses within a sentence, or production of a wrong target word. Ultimately, 3030 recordings were statistically analyzed.

2.3. Statistical analysis

The statistical analysis was conducted using the lme4 [24] and lmerTest [25] packages in R [26]. Two linear mixed-effects models were fitted for each place of articulation, i.e., alveolar and bilabial. The first model had *normalized VOT* as dependent variable. The fixed factors were *type* (lenis vs fortis), *age* (young vs. old), and *region* (rural vs. urban), including interaction terms. Random intercepts were added for *word* and *speaker* to account for individual differences. The second model was fitted for the dependent variable *normalized closure duration*. The fixed and random factors were identical to the first model.

To obtain the results of the linear mixed-effects models, a Type II ANOVA for each model was calculated using the R package car [27], which provides a readable representation of each factor and interaction. In case any of the interactions turned out to be significant, pairwise comparisons using Tukey's test were made using the R package emmeans [28].

3. Results

3.1. Alveolar plosives

The mixed-effects model for the VOT values in alveolar plosives revealed a significant main effect of *type* as well as significant interactions between *type* and *age*, *type* and *region*, *age* and *region* and the three-way interaction between *type*, *age*, and *region*. The model output can be seen in Table 1.

Table 1: *Statistical model output for the dependent variable VOT at the alveolar place of articulation.*

	Chisq	**Df**	**Pr(>Chisq)**
type	5.3359	1	.0208910*
age	2.3212	1	.1276253
region	0.3437	1	.5577177
type:age	11.5939	1	.0006617**
type:region	10.9257	1	.0009484**
age:region	4.5751	1	.0324405*
type:age:region	38.3599	1	5.883e-10**

Generally, the fortis plosives were produced with longer VOT values than the lenis plosives. Because of the significant interactions, pairwise comparisons between age groups were calculated separately for each region for both lenis and fortis plosives. Results revealed that only the fortis plosives in the urban group turned out to be significantly different between age groups, with younger speakers producing significantly longer VOT values (z=-3.932, p=.0001). This effect is also visualized in Figure 1.

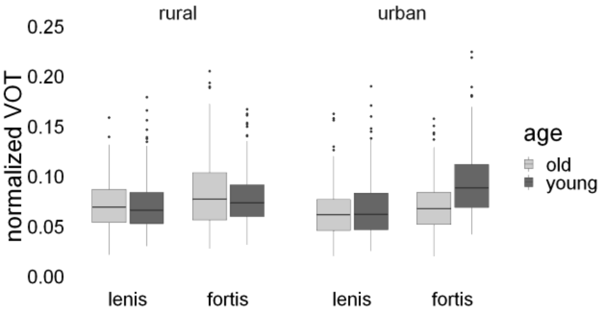

Figure 1: *normalized VOT durations (y axis) for alveolar lenis and fortis plosives (x axis); rural areas on the left, urban areas on the right; older speakers in light grey, younger speakers in dark grey.*

The model for the closure durations revealed significant main effects of *type*, *age*, and *region* as well as the interactions between *type* and *age*, *age* and *region*, and the three-way interaction between *type*, *age*, and *region*. The model output is presented in Table 2.

Table 2: *Statistical model output for the dependent variable closure duration at the alveolar place of articulation.*

	Chisq	Df	Pr(>Chisq)
type	31.5063	1	1.988e-08**
age	6.6786	1	.009758**
region	16.8930	1	3.955e-05**
type:age	23.2764	1	1.403e-06**
type:region	0.5034	1	.478000
age:region	6.7204	1	.009531**
type:age:region	9.0347	1	.002649**

Overall, closure durations are significantly longer in fortis compared to lenis plosives, as can be seen in Figure 2. This effect is larger than the difference in VOT durations, implying that closure duration remains the primary difference between lenis and fortis plosives. Due to the significant interactions, pairwise comparisons were calculated similar to those of the first model.

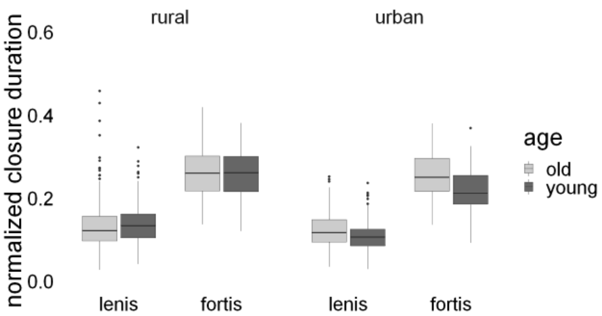

Figure 2: *normalized closure durations (y axis) for alveolar lenis and fortis plosives (x axis); rural areas on the left, urban areas on the right; older speakers in light grey, younger speakers in dark grey.*

In this case, the age-related difference in both lenis (z=1.983, p=.0473) and fortis plosives (z=4.842, p<.0001) turned out to be significant in the urban group, with younger speakers producing shorter closure durations. This effect turned out to be stronger for the fortis plosives, which is also visualized in Figure 2.

3.2. Bilabial plosives

The model for the VOT values of the bilabial plosives revealed significant main effects of *type*, *age*, and *region*, as well as the interaction between *type* and *age*.

Table 3: *Statistical model output for the dependent variable VOT at the bilabial place of articulation.*

	Chisq	Df	Pr(>Chisq)
type	7.5428	1	.0060252**
age	13.5880	1	.0002276**
region	7.3654	1	.0066491**
type:age	10.5176	1	.0011824**
type:region	0.1181	1	.7311195
age:region	2.3810	1	.1228160
type:age:region	0.0283	1	.8665078

Again, the VOT values are longer in fortis compared to lenis plosives, although this effect seems to be stronger here compared to the alveolars, as indicated by Figure 3. Because of the significant interaction, pairwise comparisons were calculated between age groups for lenis and fortis plosives separately. The results show that only in fortis plosives is the difference between younger and older speakers significant (z=-2.110, p=.0349). Still, it is obvious that this effect is stronger in the rural group, which is unexpected.

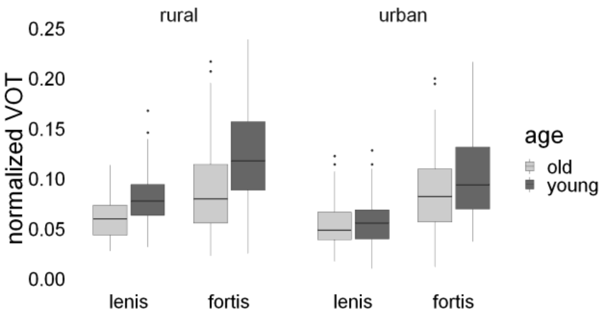

Figure 3: *normalized VOT durations (y axis) for bilabial lenis and fortis plosives (x axis); rural areas on the left, urban areas on the right; older speakers in light grey, younger speakers in dark grey.*

Moving on to the closure durations of the bilabial plosives, the mixed-effects model revealed significant main effects of *type*, *age*, and *region* as well as the interactions between *type* and *age*, *type* and *region*, and the three-way interaction between *type*, *age*, and *region*. The model output is presented in Table 4.

As expected, the closure durations for the fortis plosives are longer than those for the lenis plosives, which is shown in Figure 4. Again, this difference is larger here than the difference in the VOT values. The pairwise comparisons between age groups for each region and for lenis and fortis separately revealed significance only for the fortis plosives in the urban group (z=2.905, p=.0037).

Table 4: *Statistical model output for the dependent variable closure duration at the bilabial place of articulation.*

	Chisq	**Df**	**Pr(>Chisq)**
type	20.9317	1	4.760e-06**
age	4.9838	1	.02559*
region	5.3347	1	.02091*
type:age	15.7594	1	7.193e-05**
type:region	63.3504	1	1.730e-15**
age:region	0.1620	1	.68730
type:age:region	4.0435	1	.04434*

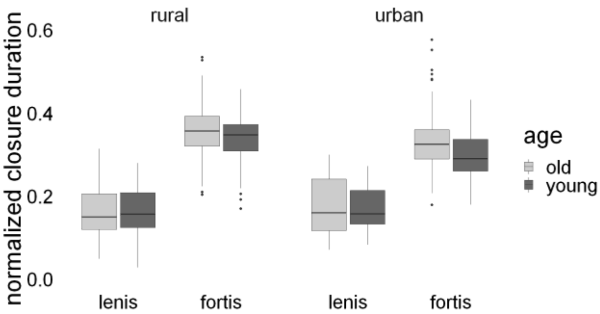

Figure 4: *normalized closure durations (y axis) for bilabial lenis and fortis plosives (x axis); rural areas on the left, urban areas on the right; older speakers in light grey, younger speakers in dark grey.*

4. Discussion

This study was conducted to investigate whether the increasing aspiration of fortis plosives in Alemannic is present in word-medial plosives in SSG as well. Specifically, the aim was to answer the questions of (1) whether younger speakers produce more aspirated fortis plosives than older speakers and (2) whether this effect would be larger in urban compared to rural areas. Furthermore, closure duration was measured to (3) investigate its relation to the VOT values.

In general, it can be confirmed that younger speakers produce longer VOT values in fortis plosives than older speakers. The effect was significant for alveolar plosives only in the urban group and in bilabial plosives for both rural and urban speakers. Therefore, the results for the alveolar plosives are in accordance with the implication that the sound change is due to the contact to GSG, while the pattern found for the bilabial plosives is slightly unexpected. One possible explanation for this effect might be that the younger speakers from the rural area produce longer VOT values in general, as shown by the lenis plosives which also show higher VOTs. Although, since this effect is much stronger for the fortis compared to the lenis plosives, it is more likely that the sound change is also emerging in the rural areas for bilabial plosives. It is also important to note that speakers of all groups produced longer VOT values for fortis compared to lenis bilabial plosives.

Interestingly, only the younger speakers from the urban area produce both longer VOT durations and shorter closure durations, while younger speakers from the rural areas do not produce significantly shorter closure durations. A possible explanation for this pattern could be that the progress of the sound change is further along in the urban area where closure duration is being reduced by younger speakers, while the younger speakers from the rural areas are only adapting the primary cue of VOT from GSG. The fact that the younger rural speakers do not alter the closure duration along with the VOT further suggests that the reduced closure durations produced by the younger urban speakers are not necessarily due to coarticulatory compensation, although this assumption requires additional investigations.

In general, the younger speakers' VOT productions are quite similar to those of the GSG speakers in [18], whose stimuli also consisted of word-medial plosives in disyllabic words and who used the same normalization method as the current study, making the results comparable.

The results provide further evidence that the current sound change in German-speaking Switzerland is most likely externally driven due to the contact to GSG. While previous research suggests that word-medial plosives seem to be affected only to a small extent in Alemannic [3], the results of the current study reveal a rather large age-related effect, partially even in rural areas, highlighting the importance to investigate not only Alemannic varieties but also SSG.

It is reasonable to assume that, due to the contact to GSG, SSG is affected first, before possible changes spread to dialectal speech. The situation in German-speaking Switzerland which comprises both a diglossia and an ongoing sound change offers thus a unique perspective on how a sound change can spread throughout different varieties.

The increase of aspiration in SSG is undoubtedly an important step of the ongoing sound change, although the change in closure duration might be even more important. Since closure duration is known as the primary cue between lenis and fortis plosives in both Alemannic and SSG, the fact that this difference is reduced in the urban group of younger speakers' productions of word-medial plosives is highly relevant to the issue of sound change, as it might be giving way to the primary cue of GSG, namely VOT. Although this situation might arise in the future, the current results show that at this point in time it is obvious that speakers of SSG (still) use closure duration as the primary cue between lenis and fortis plosives.

5. Conclusions and outlook

In conclusion, the current study highlights the importance of investigating the standard variety when dealing with a sound change in progress that is evolving in a dialectal variety. Although it has been assumed that SSG is influenced by the respective dialectal features of Alemannic, which is probably still true for many features, it has now been shown that younger speakers seem to be influenced to an increasing amount by GSG speakers. In particular, younger speakers use VOT as a distinction between lenis and fortis plosives more than older speakers. Perhaps most importantly, younger speakers from urban areas, who are assumed to have a higher amount of contact to GSG speakers, additionally produce shorter closure durations, i.e., the primary cue in SSG, in fortis plosives.

For future investigations, the most important step would now be to conduct perceptual studies for both Alemannic and SSG including VOT as well as closure duration in order to investigate their respective status in speech perception.

6. Acknowledgements

This study was funded by the Swiss National Science Foundation (SNSF), grant Nr. 164377 and 190005.

7. References

[1] H. Schifferle, “Zunehmende Behauchung. Aspirierte Plosive im modernen Schweizerdeutsch,” In: H. Christen et al. (eds.) *Alemannische Dialektologie: Wege in die Zukunft*, Steiner, pp. 43–55, 2010.

[2] R. Ladd and S. Schmid, “Obstruent Voicing Effects on F0, but without Voicing: Phonetic Correlates of Swiss German Lenis, Fortis, and Aspirated Stops,” *Journal of Phonetics*, vol. 71, pp. 229–248, 2018.

[3] F. Zebe, C. Watter, and S. Schmid, “Increasing aspiration of Swiss German plosives: a sound change in progress?,” accepted.

[4] W. Labov, *Principles of language change: Internal factors*. Blackwell, 1994.

[5] G. Bailey., T. Wikle, J. Tillery, and L. Sand, “The apparent time construct,” *Language Variation and Change*, vol. 3, pp. 241–264, 1991.

[6] C. A. Ferguson, “Diglossia,” *Word*, vol. 15, no. 2, pp. 325–40, 1959.

[7] I. Hove, *Die Aussprache der Standardsprache in der deutschen Schweiz*, Berlin & Boston: Max Niemeyer, 2002.

[8] U. Zihlmann, “Vowel and Consonant Length in Four Alemannic Dialects and Their Influence on the Respective Varieties of Swiss Standard German,” *Wiener Linguistische Gazette*, vol. 86, pp. 1–46, 2020.

[9] U. Willi, *Die segmentale Dauer als phonetischer Parameter von 'fortis' und 'lenis' bei Plosiven im Zürichdeutschen: Eine akustische und perzeptorische Untersuchung*, Franz Steiner, 1996.

[10] A. Kraehenmann, “Swiss German Stops: Geminates All over the Word,” *Phonology*, vol. 18, no. 1, pp. 109–145, 2001.

[11] J. Fleischer and S. Schmid, “Zurich German,” *Journal of the International Phonetic Association*, vol. 36, no. 2, pp. 243–253, 2006.

[12] A. Kraehenmann and A. Lahiri, “Duration Differences in the Articulation and Acoustics of Swiss German Word-Initial Geminate and Singleton Stops,” *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4446–4455, 2008.

[13] F. Zebe, “Vowel and consonant quantity in two Swiss German dialects and their corresponding varieties of Standard German: effects of region, age, and tempo,” accepted.

[14] M. Jessen, *Phonetics and phonology of tense and lax obstruents in German*, John Benjamins Publishing, 1998.

[15] E. M. Luef, “Development of voice onset time in an ongoing phonetic differentiation in Austrian German plosives: Reversing a near-merger,” *Zeitschrift für Sprachwissenschaft*, vol. 39, no. 1, pp. 79–101, 2020.

[16] F. Kleber, “VOT or quantity: What matters more for the voicing contrast in German regional varieties? Results from apparent-time analyses,” *Journal of Phonetics*, vol. 71, pp. 468–486, 2018.

[17] B. S. Phillips, “Word frequency and the actuation of sound change,” Language, 60, pp. 320–342, 1984.

[18] J. Harrington, F. Kleber, and U. Reubold, 2012. “The production and perception of coarticulation in two types of sound changes in progress,” In: S. Fuchs, M. Weirich, D. Pape, & P. Perrier (eds.) *Speech planning and Dynamics*, Peter Lang, pp. 39–62, 2012.

[19] V. Bukmaier, J. Harrington, and F. Kleber, “An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: Evidence for a gradient sound change,” *Frontiers in Psychology*, vol. 5, 2014.

[20] M. Jochim and F. Kleber, “Fast-Speech-Induced Hypoarticulation Does Not Considerably Affect the Diachronic Reversal of Complementary Length in Central Bavarian,” *Language and Speech*, 2022.

[21] C. Draxler and K. Jänsch, “SpeechRecorder – a Universal Platform Independent Multi-Channel Audio Recording Software,” *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, European Language Resources Association (ELRA), 2004.

[22] R. Winkelmann and G. Raess, *Introducing a Web Application for Labeling, Visualizing Speech and Correcting Derived Speech Signals*, 2014.

[23] R. Winkelmann, K. Jaensch, S. Cassidy, and J. Harrington, *emuR: Main Package of the EMU Speech Database Management System*, 2021.

[24] D. Bates, M. Mächler, B. Bolker, and S. Walker, 2015. “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[25] A. Kuznetsova., P. B. Brockhoff, and R. H. B. Christensen, “lmerTest Package: Tests in Linear Mixed Effects Models,” *Journal of Statistical Software* 82(13), 1–26, 2017.

[26] RStudio Team, *RStudio: Integrated Development for R. RStudio*, PBC, Boston, 2020.

[27] J. Fox and S. Weisberg, *An R companion to applied regression*, Thousand Oaks, CA: Sage publications, 2018.

[28] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2021.