EFFECTS OF COGNITIVE LOAD ON VOWEL PRODUCTION IN BILINGUAL SPEAKERS OF HIGH AND LOW GERMAN

Marina Frank, Marina Rohloff, Jörg Peters

Carl von Ossietzky University Oldenburg
{marina.frank, marina.rohloff1, joerg.peters}@uni-oldenburg.de

ABSTRACT

Cognitive load from speaking a foreign language has an influence on vocal measures such as pitch. Cognitive load induced by increasing task difficulty was found to affect the vowel space in some studies, however, the results are inconclusive. This study examines the influence of increased cognitive load from speaking a less dominant language on vowel formants in 95 bilingual speakers of High German and Low German. F1 and F2 values and vowel durations were extracted from the vowels /a/, /i/, and /u/ in stressed syllables in a route description task. We found gender-dependent differences related to language dominance for F1 and vowel duration of /a/ and for F2 of /i/. The other results point more towards language-specific differences. We conclude that cognitive load from speaking a less dominant language has some effect on vowel realizations, however, language-specific differences prevail even for very similar languages.

Keywords: Vowels, cognitive load, bilingualism, High German, Low German.

1. INTRODUCTION

Various acoustic measures are related to speech under increased cognitive load [1]–[4]. Cognitive load induced by task difficulty has been associated with an increase in disfluencies and filled pauses [5]. An increase of f0 has been reported in multiple studies, with a steeper increase for real-life stress (e.g., airplane crash) than for laboratory induced stress [6]–[11]. Additionally, acoustic measures related to the glottal signal, such as jitter and shimmer, are reduced under increased cognitive load [12]. Cepstral peak prominence has also been correlated with speech production under increased cognitive load [13].

Previous studies on the influence of cognitive load on vowel formant frequencies show contradictory results. For real-life stress in airplane incidences, [6] found higher F1 values for one pilot while [10] found no significant changes in F1 for the pilot but a significant increase in F1 for /i/ and a decrease for /u/ for the copilot. For F2, the pilot showed an increase for /a/, /e/, and /ɛ/, while the co-pilot did not show any significant differences. In studies like these, the sample size is naturally very small, i.e., only one or two pilots are investigated. In a flight simulator study with 13 subjects, [14] found increased F1 values and decreased F2 values for front vowels and increased F1 and F2 values for back vowels under cognitive load, i.e., more centralized vowels.

For laboratory induced stress, [15] found no significant differences for F1 and F2 in a dual-task condition compared to the control condition. In a Stroop test analysis of one speaker, [10] found a decrease in F1 for /u/ and for /ø/ and increased F1 and F2 values for /ɛ/ with increasing cognitive load. In their analysis of a larger Stroop test database, [16] found a decrease in F2 for /ao/; however, they did not find an overall smaller vowel space for the higher cognitive load condition. With increased cognitive load, vowels showed longer durations, potentially because subjects spoke more slowly [16]. Cognitive load was also found to affect the vowel space in anxiety-denying women (centralization), i.e., these effects varied depending on speaker sex and coping style [17]. In three different cognitive load conditions, [18] found increased F1 and F2 values compared to a control condition. In conclusion, both centralization and peripheralization of the vowel space have been observed with regard to an increase in cognitive load due to an increase in task difficulty.

It is not clear whether effects of cognitive load on vowel formants also occur in the less dominant language of bilingual speakers. Cognitive load from speaking a foreign language is reported to have an influence on f0 level and f0 dispersion ([19], [20] (English/Finnish), [21] (High/Low German)). These results are in line with the results regarding cognitive load from task difficulty. There is ample research on the influence of the vowel system of the more dominant language on the vowel system of the less dominant language (e.g., [22]–[24]). However, there have been no studies on the effect of cognitive load caused by speaking the less dominant language on vowel formants. In this case, vowel formant analysis is difficult because the acoustic characteristics of vowels often differ between two languages. For this reason, we focus on two closely related languages in this paper.

We report results of a study that examines the influence of cognitive load on vowel formants and vowel duration in bilingual speakers of High German and Low German. Low German is a regional language spoken in northern Germany that nowadays is no longer acquired monolingually, but always together with High German. Since the use of Low German is declining in younger speakers [25], we expect that they are more dominant in High German than older speakers. We expect that these speakers exhibit larger differences in F1 and F2 between the two languages than more balanced bilinguals, i.e., the middle-aged and older speakers. We also expect the speakers who are more dominant in High German to have longer vowel durations in Low German than the more balanced bilingual speakers.

2. METHOD

2.1. Participants and data collection

A total of 95 participants, aged 15 to 88 years, took part in this study. All subjects (47 female, 48 male) grew up in the municipality of Krummhörn in East Frisia, in the northwest of the federal state of Lower Saxony in Germany. The subjects were native speakers of East Frisian Low German (LG) and the regional standard variety of High German (HG). We had to exclude 4 participants because of missing data, retaining 91 subjects for analysis.

All subjects verbally completed a questionnaire consisting of 55 questions on socio-demographics as well as on age of acquisition, frequency of language use, and self-attributed language proficiency (adopted from [25–27]). Of all questions, 28 were used to calculate global language scores for each subject and language (based on [26] and [27]). To obtain the language dominance score used in our analysis, we subtracted the HG values from the LG values. Negative values indicate a dominance of HG over LG (= HG dominant), positive values a dominance of LG over HG (= LG dominant). A value around 0 indicates balanced bilingualism. Possible values range from -174 to +174; the values in our sample range from -119 to +58. The younger speakers show the highest dominance of HG while the older speakers exhibit balanced dominance scores. For a more thorough analysis of the correlation between age and dominance score in our subjects see [28].

Speech recordings were collected using a head-mounted omnidirectional microphone (DPA 4066) and a portable digital recorder (Tascam DR-100 MKIII). The recordings were digitized at 48 kHz sampling rate with 24 bits/sample quantization.

2.2. Procedure and data selection

During the experiment, subjects completed seven tasks in each language. The order of the languages was randomized per subject, the order of the tasks was kept stable. One of these tasks was giving a route description, for which we varied the level of task difficulty to induce different levels of cognitive load. In the more difficult condition, the participants received a map without any labels or landmarks. In the easier condition, labels and landmarks were added to the map. The routes were different in each condition.

We manually segmented the point vowels /i/, /u/, and /a/ in stressed syllables in the route description tasks. We excluded vowels in the context before /r/, as post-vocalic /r/ is vocalized in LG and in HG resulting in a diphthong [29][30]. For /i/ and /u/, we do not expect any language-specific spectral differences [29][30]. For /a/, we expect a central open realization for HG [29] and a more retracted and slightly rounded realization for LG [30]. Regarding vowel duration, we do not expect language-specific differences for all three vowels.

The duration of the route description tasks ranged from 15.87 s to 176.31 s ($\bar{x}$ = 68.93 s). In total, the sample contained 2,470 /a/ tokens, 2,343 /i/ tokens, and 1,409 /u/ tokens.

2.3. Acoustic analysis

F1 and F2 values were measured at vowel midpoint using Praat [31]. We calculated five formants using the Burg LPC method. The formant ceiling was set to 5000 Hz for male speakers and to 5500 Hz for female speakers. Window length was 25 ms and pre-emphasis from 50 Hz. Additionally, we extracted the duration for each vowel. All recordings were downsampled to 16 kHz before analysis. During the inspection of the automatic formant extraction, we found obvious measurement errors. To reduce these, we excluded the 5% outermost values for each vowel and each formant, leaving 2,235 /a/ tokens (HG: 1,207; LG: 1,263), 2,115 /i/ tokens (HG: 1,294; LG: 1,049), and 1,286 /u/ tokens (HG: 370; LG: 1,039) for analysis.

2.4. Statistical analysis

We fitted LMMs in R [32] using the glmmTMB package [33]. The dependent variables were F1, F2, and Duration of each vowel. As fixed effects, we added the within-subject effect Language (HG vs. LG) and the between-subjects effects Dominance Score and Gender (female vs. male) to the model as well as all possible interactions. We used random intercepts for Subjects as well as by-subject random slopes for Language. Since the residuals for the dependent variables were not normally distributed, we used the Gamma distribution and log-link function in our models. Interaction results are reported using the emtrends function from the emmeans package [34]. The level of significance was set to p = 0.05. Figures were created using tidyverse packages [35].

We included the main effects for Gender and Dominance Score only as control variables, which is why we do not report these results below. The factor Task Difficulty (easy vs. complex) was left out of the models because the effects on vowel production were inconsistent and difficult to interpret. This could probably be because the difficulty levels did not differ sufficiently from each other to reliably affect vowel realizations.

3. RESULTS

3.1. F1

Figure 1 displays F1 values for each vowel as a function of Dominance Score per Gender.

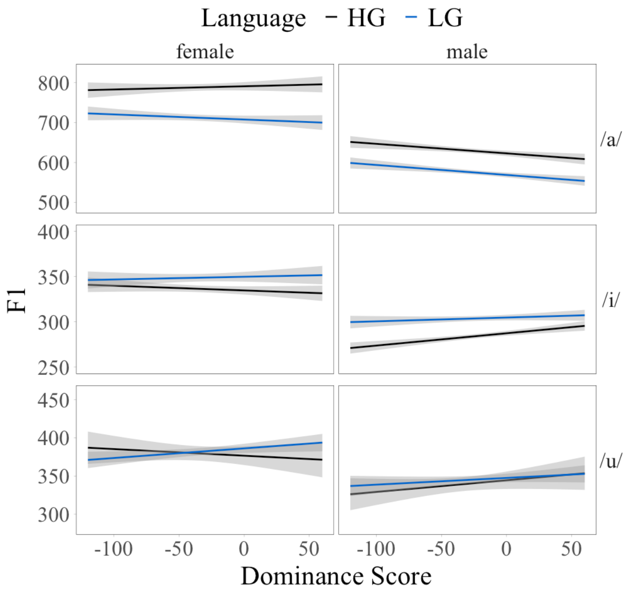

Figure 1: F1 values. Shading indicates 95% confidence intervals.

For /a/, statistical analysis revealed significantly lower F1 values for LG compared to HG ($\chi^2 = 185.691$, $p < .001$).

For /i/, statistical analysis revealed significantly higher F1 values for LG compared to HG ($\chi^2 = 63.394$, $p < .001$). In addition, we found a three-way interaction between Language, Gender, and Dominance Score. Contrast analysis revealed a significantly steeper, positive slope for HG compared to LG for male speakers (estimate = -0.0005, $SE = 0.0002$, z-ratio = -2.051, $p < .05$), suggesting that LG dominant speakers exhibit smaller differences in F1 between the two languages than HG dominant speakers.

For /u/, statistical analysis revealed significantly higher F1 values for LG compared to HG ($\chi^2 = 4.484$, $p < .05$).

3.2. F2

Figure 2 displays F2 values for each vowel as a function of Dominance Score per Gender.

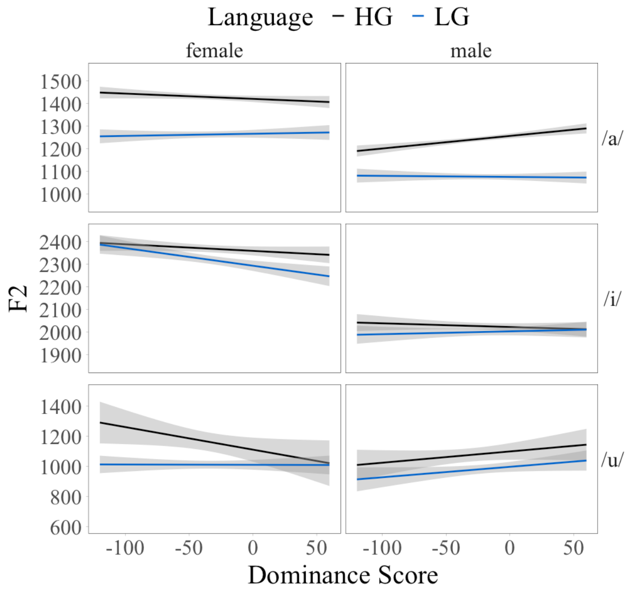

Figure 2: F2 values. Shading indicates 95% confidence intervals.

For /a/, statistical analysis revealed significantly lower F2 values for LG compared to HG ($\chi^2 = 440.83$, $p < .001$). Additionally, we found a three-way interaction between Language, Gender, and Dominance Score. Contrast analysis revealed a significantly steeper, positive slope for HG compared to a slightly negative slope for LG for male speakers (estimate = -0.0005, $SE = 0.0002$, z-ratio = -2.299, $p < .05$), suggesting that LG dominant speakers exhibit larger differences in F2 between the two languages than HG dominant speakers.

For /i/ and /u/, statistical analysis revealed significantly lower F2 values for LG compared to HG (/i/: $\chi^2 = 12.766$, $p < .001$; /u/: $\chi^2 = 16.784$, $p < .001$).

3.3. Duration

Figure 3 displays Duration values for each vowel as a function of Dominance Score per Gender.

For /a/, statistical analysis revealed significantly longer vowels for LG compared to HG ($\chi^2 = 29.353$, $p < .001$). Additionally, we found a three-way interaction between Language, Gender, and Dominance Score. Contrast analysis revealed a significantly steeper, positive slope for HG compared to a negative slope for LG for female speakers (estimate = -0.0013, $SE = 0.0006$, z-ratio = -2.340, $p < .05$), suggesting that LG dominant speakers exhibit a smaller difference in Duration between the two languages than HG dominant speakers.

For /i/ and /u/, statistical analysis revealed significantly longer vowels for LG compared to HG (/i/: $\chi^2 = 58.117$, $p < .001$; /u/: $\chi^2 = 17.348$, $p < .001$).

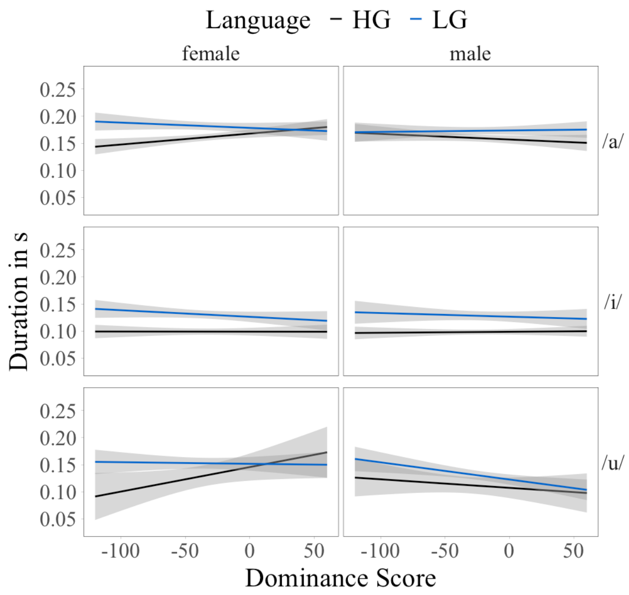

Figure 3: Duration values. Shading indicates 95% confidence intervals.

4. DISCUSSION

In this study, formant values and vowel durations of the point vowels /a/, /i/, and /u/ were analyzed from route descriptions provided by 91 bilingual speakers. Our hypothesis was that HG dominant speakers, i.e., mostly younger speakers, would show larger differences between LG and HG than LG dominant speakers, i.e., middle-aged and older speakers, as for the former the use of the less dominant language, LG, is associated with increased cognitive load. We conducted nine analyses, including F1, F2, and vowel duration for each of the three vowels, and found that language dominance had a partial effect on male speakers' F1 values for /i/ and F2 values for /a/, as well as on female speakers' durational values for /a/.

The lower F1 and F2 values for /a/ in LG suggest that /a/ is pronounced as a more closed back vowel in LG compared to HG. These results can directly be attributed to language-specific differences [30]. Additionally, a three-way interaction between Language, Gender, and Dominance Score showed a larger difference in F2 for LG dominant male speakers than for HG dominant male speakers. This effect is the contrary of what we hypothesized. Moreover, the difference can be attributed to a change in F2 in HG while the F2 values in LG remain relatively stable across speakers with varying language dominance.

For /i/, we found higher F1 values and lower F2 values in LG compared to HG. In combination, these values indicate a less peripheral production of /i/ in LG. Furthermore, a three-way interaction between Language, Gender, and Dominance Score showed a smaller difference in F1 for LG dominant male speakers than for HG dominant male speakers, as we hypothesized. However, the difference can be attributed to a change in F1 in HG while the F1 values in LG remain relatively stable across speakers independent of their dominance score.

The higher F1 values for /u/ in LG compared to HG indicate a more open production of /u/ in LG. In contrast, lower F2 values in LG indicate a further back production of /u/ in LG. As we find no interactions with language dominance for /u/, we cannot decide whether these effects arise from increased cognitive load in speaking the less dominant language or from language-specific differences.

Regarding the durational values, we found longer vowel durations in LG than in HG for /a/, /i/, and /u/. These results are in line with [16], who found it an important measure for increased cognitive load. For /a/, a three-way interaction between Language, Gender, and Dominance Score showed converging durational values for LG dominant female speakers. However, for the most part we cannot decide whether these are language-specific or cognitive load-related differences.

In only three cases, our results include a partial effect of language dominance; and even these cases are contradictory. Considering the results at hand, the question arises whether vowel formants and durations are suitable indicators for cognitive load in bilingual speech. The inconsistent results of our study are in line with previous research. One possible explanation would be that different coping styles as reported in [17] have led to opposing effects that have largely cancelled each other out.

5. CONCLUSION

Our study revealed only a few effects of increased cognitive load from speaking a less dominant language on vowel production. Most of our findings could also be attributed to language-specific differences. Nevertheless, this study adds to the existing literature by investigating cognitive load in spontaneous speech of bilingual speakers.

6. ACKNOWLEDGMENTS

The research presented here was funded by the German Research Foundation (DFG), grant number PE 793/3–1. We thank our participants for their time and effort, our student assistants for help with data cleaning and analysis, and our recording assistants for their help with data collection.

7. REFERENCES

[1] Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., Winter, A. S. 2013. Vocal Indices of Stress: A Review. *J. Voice* 27, 390.e21–390.e29.

[2] Kirchhübel, C., Howard, D. M., Stedmon, A. W. 2011. Acoustic correlates of speech when under stress: Research, methods and future directions. *Int. J. Speech Lang. Law* 18, 75–98.

[3] Moore, R. K. (ed.) 1996. Speech under Stress [Special Issue]. *Speech Commun.* 20.

[4] Van Puyvelde, M., Neyt, X., McGlone, F., Pattyn, N. 2018. Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance. *Front. Psychol.* 9, 1994.

[5] Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., Wittig, F. 2001. Recognizing Time Pressure and Cognitive Load on the Basis of Speech: An Experimental Study. *User Modeling: Proceedings of the Eighth International Conference*, 24–33.

[6] Benson, P. 1995. Analysis of the Acoustic Correlates of Stress from an Operational Aviation Emergency. *ESCA-NATO Workshop on Speech under Stress*, 61–64.

[7] Brenner, M., Shipp, T., Doherty, E. T., Morrissey, P. 1985. Voice measures of psychological stress: laboratory and field data. In: Titze, I. R., Scherer, R. C. (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*. Denver Center for Performing Arts, 239–248.

[8] Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., Leino, T. 2011. Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Appl. Ergon.* 42, 348–357.

[9] Jessen, M. 2006. *Einfluss von Stress auf Sprache und Stimme: Unter besonderer Berücksichtigung polizeidienstlicher Anforderungen*. Schulz-Kirchner.

[10] Ruiz, R., Absil, E., Harmegnies, B., Legros, C., Poch, D. 1996. Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Commun.* 20, 111–129.

[11] Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., Bänziger, T. 2002. Acoustic correlates of task load and stress. *7th ICSLP 2002 Denver Colo. USA Sept. 16-20 2002*, 2017–2020.

[12] Boyer, S., Paubel, P.-V., Ruiz, R., El Yagoubi, R., Daurat, A. 2018. Human Voice as a Measure of Mental Load Level. *JSLHR* 61(11), 2722–2734.

[13] Yap, T. F., Epps, J., Ambikairajah, E., Choi, E. H. C. 2015. Voice source under cognitive load: Effects and classification. *Speech Commun.* 72, 74–95.

[14] Huttunen, K. H., Keränen, H. I., Pääkkönen, R. J., Päivikki Eskelinen-Rönkä, R., Leino, T. K. 2011. Effect of cognitive load on articulation rate and formant frequencies during simulator flights. *J. Acoust. Soc. Am.* 129, 1580–1593.

[15] Lively, S. E., Pisoni, D. B., Van Summers, W., Bernacki, R. H. 1993. Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences *J. Acoust. Soc. Am.* 93, 2962–2973.

[16] Yap, T. F., Epps, J., Ambikairajah, E., Choi, E. H. C. 2011. Formant Frequencies under Cognitive Load: Effects and Classification. *EURASIP J. Adv. Signal Process.* 2011, 219253.

[17] Tolkmitt, F. J., Scherer, K. J. 1986. Effect of Experimentally Induced Stress on Vocal Parameters. *J. Exp. Psychol. Hum. Percept. Perform.* 12(3), 302–313.

[18] Meinerz, C. 2010. *Effekte von Stress auf Stimme und Sprechen: Eine phonetische Untersuchung auf der Grundlage ausgewählter akustischer und sprechdynamischer Parameter unter Berücksichtigung verschiedener Stressklassen*. Books on Demand.

[19] Järvinen, K., Laukkanen, A.-M. 2015. Vocal loading in speaking a foreign language. *Folia Phoniatr. Logop.* 67, 1–7.

[20] Järvinen, K., Laukkanen, A.-M., Aaltonen, O. 2013. Speaking a foreign language and its effect on F0. *Logoped. Phoniatr. Vocol.* 38, 47–51.

[21] Peters, J., Frank, M., Rohloff, M. 2020. Pitch range variation in High German (L1) and Low German (L2). *Proceedings of the 10th International Conference on Speech Prosody*, 650–654.

[22] Mayr, R., Morris, J., Mennen, I., Williams, D. 2017. Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English. *Int. J. Biling.* 21, 245–267.

[23] Schoormann, H. E., Heeringa, W., Peters, J. 2019. Standard German vowel productions by monolingual and trilingual speakers. *Int. J. Biling.* 23, 138–156.

[24] Yang, J. 2021. Vowel development in young Mandarin-English bilingual children. *Phonetica* 78, 241–272.

[25] Adler, A., Ehlers, C., Goltz, R., Kleene, A., Plewnia, A. 2016. *Status und Gebrauch des Niederdeutschen 2016: Erste Ergebnisse einer repräsentativen Erhebung*. Institut für Deutsche Sprache.

[26] Birdsong, D., Gertken, L. M., Amengual, M. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. sites.la.utexas.edu/bilingual/.

[27] Poarch, G. J., Vanhove, J., Berthele, R. 2019. The effect of bidialectalism on executive function. *Int. J. Biling.* 23, 612–628.

[28] Peters, J., Frank, M., Rohloff, M. 2023. Vocal fold vibratory patterns in bilingual speakers of Low and High German. *Proceedings of the 20th International Congress of Phonetic Sciences.*

[29] Kleiner, S., Knöbl, R., Dudenredaktion, Institut für Deutsche Sprache (eds.). 2015. *Duden: Das Aussprachewörterbuch*. Dudenverlag & Institut für Deutsche Sprache.

[30] Lücht, W. 2016. *Ostfriesische Grammatik*. Ostfriesische Landschaft.

[31] Boersma, P., Weenink, D. 2022. Praat: Doing Phonetics by Computer [Computer program]. V. 6.2.14., retrieved 24 May 2022. www.fon.hum.uva.nl/praat/.

[32] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. www.r-project.org

[33] Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., Bolker, B. M. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* 9, 378–400.

[34] Lenth, R. V. 2022. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.5. https://cran.r-project.org/package=emmeans.

[35] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. 2019. Welcome to the tidyverse. *JOSS* 4, 1686.